

ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES

RELATED APPLICATIONS

- This application is a continuation of U.S. application Serial No.
- 5 08/629,822, filed April 10, 1996. This application is related to U.S. application Serial No. 08/759,558, now U.S. Patent No. 5,288,625 and to U.S. application Serial No. 08/375,271, filed 1/19/95, which is a continuation of U.S. application Serial No. 08/080,097, filed 6/23/93 which is a continuation of U.S. application Serial No. 07/892,487, filed
- 10 6/3/92, which is a continuation of U.S. application Serial No. 07/521,073, filed 5/9/90.

The subject matter of each of the above-listed U.S. applications is incorporated in its entirety by reference thereto.

FIELD OF THE INVENTION

- 15 The present invention relates to methods for preparing cell lines that contain artificial chromosomes, to methods for isolation of the artificial chromosomes, targeted insertion of heterologous DNA into the chromosomes, isolation of the chromosomes, and delivery of the chromosomes to selected cells and tissues. Also provided are cell lines for use in
- 20 the methods, and cell lines and chromosomes produced by the methods.

BACKGROUND OF THE INVENTION

- Several viral vectors, non-viral, and physical delivery systems for gene therapy have been developed [see, e.g., Mitani et al. (1993) Trends Biotech. 11:162-166]. The presently available systems, however, have
- 25 numerous limitations, particularly where persistent, stable, or controlled gene expression is required. These limitations include: (1) size limitations because there is a limit, generally on order of about ten kilobases [kB], at most, to the size of the DNA insert [gene] that can be accepted by viral vectors, whereas a number of mammalian genes of possible therapeutic
- 30 importance are well above this limit, especially if all control elements are included; (2) the inability to specifically target integration so that random

integration is required which carries a risk of disrupting vital genes or cancer suppressor genes; (3) the expression of randomly integrated therapeutic genes may be affected by the functional compartmentalization in the nucleus and are affected by chromatin-based position effects; (4) 5 the copy number and consequently the expression of a given gene to be integrated into the genome cannot be controlled. Thus, improvements in gene delivery and stable expression systems are needed [see, e.g., Mulligan (1993) Science 260:926-932].

In addition, safe and effective gene therapy methods and vectors 10 should have numerous features that are not assured by the presently available systems. For example, a safe vector should not contain DNA elements that can promote unwanted changes by recombination or mutation in the host genetic material, should not have the potential to initiate deleterious effects in cells, tissues, or organisms carrying the 15 vector, and should not interfere with genomic functions. In addition, it would be advantageous for the vector to be non-integrative, or designed for site-specific integration. Also, the copy number of therapeutic gene(s) carried by the vector should be controlled and stable, the vector should secure the independent and controlled function of the introduced gene(s); 20 and the vector should accept large (up to Mb size) inserts and ensure the functional stability of the insert.

The limitations of existing gene delivery technologies, however, argue for the development of alternative vector systems suitable for transferring large [up to Mb size or larger] genes and gene complexes 25 together with regulatory elements that will provide a safe, controlled, and persistent expression of the therapeutic genetic material.

At the present time, none of the available vectors fulfill these requirements. Some of these characteristics, however, are possessed by chromosomes. Thus, an artificial chromosome would be an ideal vector 30 for gene therapy, as well as for production of gene products that require

09096648 "061298

coordination of expression of numerous genes or that are encoded by large genes, and other uses. Artificial chromosomes for expression of heterologous genes in yeast are available, but construction of a mammalian artificial chromosome has not been achieved. Such

5 construction has been hindered by the lack of an isolated, functional, mammalian centromere and uncertainty regarding the requisites for its production and stable replication. Unlike in yeast, there are no selectable genes in close proximity to a mammalian centromere, and the presence of long runs of highly repetitive pericentric heterochromatic DNA makes the

10 isolation of a mammalian centromere using presently available methods, such as chromosome walking, virtually impossible. Other strategies are required for production of mammalian artificial chromosomes, and some have been developed. For example, U.S. Patent No. 5,288,625 provides a cell line that contains an artificial chromosome, a minichromosome, that

15 is about 20 to 30 megabases. Methods provided for isolation of these chromosomes, however, provide preparations of only about 10-20% purity. Thus, development of alternative artificial chromosomes and perfection of isolation methods as well as development of more versatile chromosomes and further characterization of the minichromosomes is

20 required to realize the potential of this technology.

Therefore, it is an object herein to provide mammalian artificial chromosomes and methods for introduction of foreign DNA into such chromosomes. It is also an object herein to provide methods for introduction of the artificial mammalian chromosome into selected cells,

25 and to provide the resulting cells, as well as transgenic animals and plants that contain the artificial chromosomes. It is also an object herein to provide methods for gene therapy and expression of gene products using artificial chromosomes. It is a further object herein to provide methods for constructing species-specific artificial chromosomes.

SUMMARY OF THE INVENTION

Mammalian artificial chromosomes [MACs] are provided. Also provided are artificial chromosomes for other higher eukaryotic species, such as insects and fish, produced using the MACS are provided herein.

- 5 Methods for generating and isolating such chromosomes. Methods using the MACs to construct artificial chromosomes from other species, such as insect and fish species are also provided. The artificial chromosomes are fully functional stable chromosomes. Two types of artificial chromosomes are provided. One type, herein referred to as SATACs
- 10 [satellite artificial chromosomes] are stable heterochromatic chromosomes, and the another type are minichromosomes based on amplification of euchromatin.

- Artificial chromosomes permit targeted integration of megabase pair size DNA fragments that contain single or multiple genes. Thus methods
- 15 using the MACs to introduce the genes into cells, animals and tissues are also provided. The artificial chromosomes with integrated heterologous DNA are to be used in methods of gene therapy, in methods of production of gene products, particularly products that require expression of multigene biosynthetic pathways, and also are intended for delivery
- 20 into the nuclei of germline cells, such as embryo-derived stem cells [ES cells] for production of transgenic animals.

- Mammalian artificial chromosomes provide extra-genomic specific integration sites for introduction of genes encoding proteins of interest and permit megabase size DNA integration so that, for example, genes
- 25 encoding an entire metabolic pathway or a very large gene, such as the cystic fibrosis [CF; ~600kb] gene, several genes, such as a series of antigens for preparation of a multivalent vaccine, can be stably introduced into a cell. Vectors for targeted introduction of such genes, including the tumor suppressor genes, such as p53, the cystic fibrosis transmembrane

regulator gene [CFTR], anti-HIV ribozymes, such as an anti-HIV gag ribozyme, into the artificial chromosomes also provided.

5 The chromosomes provided herein are generated by introducing heterologous DNA that includes DNA encoding a selectable marker into cells, preferably a stable cell line, growing the cells under selective conditions, and identifying from among the resulting clones those that include chromosomes with more than one centromere or that have chromosomes that are fragments of chromosomes that had more than one centromere. The amplification that produces the additional
10 centromere occurs in cells that contain chromosomes in which the heterologous DNA has integrated near the centromere in the pericentric region of the chromosome. The selected clonal cells are then used to generate artificial chromosomes.

15 In preferred embodiments, the DNA with the selectable marker that is introduced includes sequences that target it to the pericentric region of the chromosome. For example, vectors, such as pTEMPUD, which includes such DNA specific for mouse satellite DNA, are provided. Also provided are derivatives of pTEMPUD that specifically target human satellite sequences. Upon integration, these vectors can induce the
20 amplification.

Artificial chromosomes are generated by culturing the cells with the dicentric chromosomes under conditions whereby the chromosome breaks to form a minichromosome and formerly dicentric chromosome. The artificial chromosomes [the SATACs] are generated, not from the
25 minichromosome fragment as, for example, in U.S. Patent No. 5,288,625, but from the fragment of the formerly dicentric chromosome.

Among the MACs provided herein are the SATACs, which are

primarily made up of repeating units of short satellite DNA and are fully heterochromatic, so that absent insertion of heterologous or foreign DNA, the chromosomes do not contain genetic information. They can thus be used as "safe" vectors for delivery of DNA to mammalian hosts because
5 they do not contain any potentially harmful genes.

In addition to MACs methods for generating euchromatic minichromosomes and the use thereof are also provided herein. Methods for generating one type of MAC the minichromosome, previously described in U.S. Patent No. 5,288,625, and the use
10 thereof for expression of heterologous DNA are provided. Cell lines containing the minichromosome and the use thereof for cell fusion are also provided.

In one embodiment, a cell line containing the mammalian minichromosome is used as recipient cells for donor DNA encoding a
15 selected gene or multiple genes. The donor DNA is linked to a second selectable marker and is targeted to and integrated into the minichromosome. The resulting chromosome is transferred by cell fusion into an appropriate recipient cell line, such as a Chinese hamster cell line [CHO]. After large scale production of the cells carrying the engineered
20 chromosome, the chromosome is isolated. In particular, metaphase chromosomes are obtained, such as by addition of colchicine, and they are purified from the cell lysate. These chromosomes are used for cloning, sequencing and for delivery of heterologous DNA into cells.

Also provided are SATACs of various sizes that are formed by
25 repeated culturing under selective conditions and subcloning of cells that contain chromosomes produced from the formerly dicentric chromosomes. These chromosomes are based on repeating units 7.5 to 10 Mb referred to herein as megareplicons, that are tandem blocks of satellite DNA flanked by heterologous non-satellite DNA. Amplification
30 produces a tandem array of identical chromosome segments [each called

an amplicon] that contain two inverted megareplicons bordered by heterologous ["foreign"] DNA. Repeated cell fusion, growth on selective medium and/or BrdU [5-bromodeoxyuridine] treatment or other genome destabilizing reagent or agent, such as ionizing radiation, including X-rays, and subcloning results in cell lines that carry stable heterochromatic or partially heterochromatic chromosomes, including a 150-200 Mb "sausage" chromosome, a 500-1000 Mb gigachromosome, a stable 250-400 Mb megachromosome and various smaller stable chromosomes derived therefrom. These chromosomes are based on these repeating units and can include heterologous DNA that is expressed.

Thus methods for producing MACs of both types are provided. These methods are applicable to any higher eukaryotic cell, including mammals, insects and plants.

The resulting chromosomes can be purified by methods provided herein to provide vectors for introduction of the heterologous DNA into selected cells for production of the gene product encoded by the heterologous DNA, for production of transgenic animals and plants or for gene therapy. Vectors for chromosome fragmentation are provided. These vectors will be used to produce an artificial chromosome that contains a megareplicon, a centromere and two telomeres and will be between about 10 Mb and about 60 Mb, preferably between about 10 Mb-15 Mb and 30 Mb. Such artificial chromosomes may be produced by other methods. Isolation of the 7.5 Mb [or 15 Mb amplicon containing two 7.5 Mb inverted repeats] or a 30 Mb multimer thereof should provide a stable chromosomal vector that can be manipulated in vitro.

In addition, methods and vectors for fragmenting the minichromosomes and SATACs are provided. Such methods and vectors can be used for in vivo generation of smaller stable artificial chromosomes. Methods for reducing the size of the MACs to generate smaller stable self-replicating artificial chromosomes are also provided.

Methods and vectors for targeting heterologous DNA into the artificial chromosomes are also provided as are methods and vectors for fragmenting the chromosomes to produce smaller but stable and self-replicating artificial chromosomes. Vectors for targeted introduction of heterologous DNA into artificial chromosomes are provided, ~~as are~~

- 5 The chromosomes are introduced into cells to produce stable transformed cell lines or cells, depending upon the source of the cells. Introduction is effected by any suitable method including, but not limited to electroporation, direct uptake, such as by calcium phosphate [see, e.g.,
10 Wigler et al. (1979) Proc. Natl. Acad. Sci. U.S.A. 76:1373-1376; and Current Protocols in Molecular Biology, Vol. 1, Wiley Inter-Science, Supplement 14, Unit 9.1.1-9.1.9 (1990)]. precipitation, uptake of isolated chromosomes by lipofection, by microcell fusion [see, EXAMPLES, see, also Lambert (1991) Proc. Natl. Acad. Sci. U.S.A. 88:5907-5911, U.S.
15 Patent No. 5,396,767] or other suitable method. The resulting cells can be used for production of proteins in the cells. The chromosomes can be isolated and used for gene delivery.

- Methods for isolation of the chromosomes based on the DNA content of the chromosomes, which differs from the authentic
20 chromosomes are provided.

- These artificial chromosomes can be used in gene therapy, gene product production systems, production of humanized organs, production of transgenic plants and animals, including invertebrates, vertebrate, reptiles and insects, any organism or device that would employ
25 chromosomal elements as information storage vehicles, and also for analysis and study of centromere function, for the production of artificial chromosome vectors that can be constructed in vitro, and for the preparation of species- specific artificial chromosomes. The artificial chromosomes can be introduced into cells using microinjection, cell
30 fusion, microcell fusion, electroporation, electrofusion, projectile

bombardment, calcium phosphate precipitation, site-specific targeting and other such methods. Cells particularly suited for use with the artificial chromosomes include, but are not limited to plant cells, particularly tomato, arabidopsis, and others, insect cells, including silk worm cells, 5 insect larvae, fish, reptiles, amphibians, arachnids, mammalian cells, embryonic stem cells, embryos and cells for use in methods of genetic therapy, such as lymphocytes that are used in methods of adoptive immunotherapy and nerve or neural cells. Thus methods of producing gene products and transgenic animals and plants are provided. Also 10 provided are the resulting transgenic animals and plants.

Exemplary cell lines that contain these chromosomes are also provided.

Methods for preparing artificial chromosomes for particular species and for cloning centromeres are also provided. In particular, a method for 15 cloning a centromere from an animal or plant by preparing a library of DNA fragments that contain the genome of the plant or animal, introducing the each of the fragments into a mammalian satellite artificial chromosome [SATAC] that contains a centromere from a different species, generally a mammal, from the selected plant or animal, generally 20 a non-mammal, and a selectable marker. The selected plant or animal is one in which the mammalian species centromere does not function. Each of the SATACs is introduced into the cells, which are grown under selective conditions, and cells with SATACs are identified. Such SATACS should contain a centromere encoded by the DNA from the library.

25 Also provided are libraries in which the relatively large fragments of DNA are contained on artificial chromosomes.

Transgenic animals, invertebrates and vertebrates, plants and insects, fish, reptiles, amphibians, arachnids and mammals are also provided. Of particular interest are transgenic animals that express genes 30 that confer resistance or reduce susceptibility to disease. Since multiple

genes can be introduced on a MAC, a series of genes encoding an antigen can be introduced, which up expression will serve to immunize [in a manner similar to a multivalent vaccine] the host animal against the diseases for which exposure to the antigens provide immunity or some protection.

5 Methods for cloning centromeres, such as mammalian centromeres, are also provided. In particular, in one embodiment, a library composed of fragments of the SATACs are cloned into YACs [yeast artificial chromosomes] that include a detectable marker, such as DNA encoding tyrosinase and then introduced into mammalian cells, such as albino mouse embryos. Mice produced from YACs that include a centromere that functions in mammals will express the detectable marker. Thus, if albino mice will be pigmented or have regions of pigmentation.

DESCRIPTION OF THE DRAWINGS

15 Figure 1 is a schematic drawing depicting formation of the MMCneo [the minichromosome] chromosome. A-G represents the successive events consistent with observed data that would lead to the formation and stabilization of the minichromosome.

Figure 2 shows a schematic summary of the manner in which the observed new chromosomes would form, and the relationships among the different *de novo* formed chromosomes. In particular, this figure shows a schematic drawing of the *de novo* chromosome formation initiated in the centromeric region of mouse chromosome 7. (A) A single E-type amplification in the centromeric region of chromosome 7 generates a neo-centromere linked to the integrated "foreign" DNA, and forms a dicentric chromosome. Multiple E-type amplification forms the λ neo-chromosome, which was derived from chromosome 7 and stabilized in a mouse-hamster hybrid cell line; (B) Specific breakage between the centromeres of a dicentric chromosome 7 generates a chromosome fragment with the neo-centromere, and a chromosome 7 with traces of heterologous DNA at the

(figure 2-3)

end_λ (C) Inverted duplication of the fragment bearing the neo-centromere results in the formation of a stable neo-minichromosome; (D) Integration

of exogenous DNA into the heterologous DNA region of the formerly dicentric chromosome 7 initiates H-type amplification, and the formation

5 of a heterochromatic arm. By capturing a euchromatic terminal segment, this new chromosome arm is stabilized in the form of the "sausage"

chromosome; (E) BrdU [5-bromodeoxyuridine], treatment and/or drug selection induce further H-type amplification, which results in the

formation of an unstable gigachromosome; (F) Repeated BrdU treatments

10 and/or drug selection induce further H-type amplification including a centromere duplication, which leads to the formation of another

heterochromatic chromosome arm. It is split off from the chromosome 7 by chromosome breakage, and by acquiring a terminal segment, the

stable megachromosome is formed.

15 Figure 3 Schematic diagram of the replicon structure and a scheme by which the megachromosome could be produced.

Figure 4 sets forth the relationships among the some of the exemplary cell lines described herein.

Figure 5 is diagram of the plasmid pTEMPUD.

20 DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. All patents and publications

25 referred to herein are incorporated by reference.

As used herein, a mammalian artificial chromosome [MAC] is piece of DNA that can stably replicate and segregate alongside endogenous chromosomes. It has the capacity to accommodate and express heterologous genes inserted therein. It is referred to as a mammalian

30 artificial chromosome because it includes an active mammalian

centromere. Plant artificial chromosomes and an insect artificial chromosomes refer to chromosomes that include plant and insect centromeres, respectively. A human specific chromosome [HAC] refers to chromosomes that include human centromeres, BUGACs refer to 5 artificial insect chromosomes, and AVACs refer to avian artificial chromosomes.

As used herein, stable maintenance of chromosomes, occurs when at least about 85%, preferably 90%, more preferably 95%, of the cells retain the chromosome. Stability is measured in the presence of selective 10 agent. Preferably these chromosomes are also maintained in the absence of a selective agent. Stable chromosomes also retain their structure during cell culturing, suffering neither intrachromosomal nor interchromosomal rearrangements.

As used herein, growth under selective conditions, means growth 15 of a cell under conditions that require expression of a selectable marker for survival.

As used herein, euchromatin and heterochromatin have their recognized meanings, euchromatin refers to DNA that contains genes, and heterochromatin refers to chromatin that has been thought to be 20 inactive. Highly repetitive DNA sequences [satellite DNA] are located in regions of centromeric heterochromatin [pericentric heterochromatin]. Constitutive heterochromatin refers to heterochromatin that contains the highly repetitive DNA and that is constitutively condensed.

As used herein, BrdU refers to 5-bromodeoxyuridine, which during 25 replication is inserted in place of thymidine. BrdU is used as mutagen; it also inhibits condensation of metaphase chromosomes during cell division.

As used herein, a dicentric chromosome is a chromosome that contains two centromeres. A multicentric chromosome contains more 30 than two centromeres.

As used herein, a formerly dicentric chromosome is a chromosome that is produced when a dicentric chromosome fragments and acquires new telomeres so that two chromosomes, each having one of the centromeres, are produced. Each of the fragments, are replicable

5 chromosomes. If one of the chromosomes undergoes amplification of euchromatic DNA to produce a full functionally chromosome that contains the heterologous DNA and primarily [at least more than 50%] euchromatin, it is a minichromosome. The remaining chromosome is a

10 formerly dicentric chromosome. If one of the chromosomes undergoes amplification, whereby heterochromatin [satellite DNA] is amplified, a euchromatic portion [or arm remains], it is referred to as a sausage chromosome. A chromosome that is substantially all heterochromatin, except for portions of heterologous DNA, is called a SATAC. Such

15 chromosomes [SATACs] can be produced from sausage chromosomes by culturing the cell containing the sausage chromosome under conditions, such as BrdU treatment and/or growth under selective conditions, that destabilize the chromosome so that a satellite artificial chromosomes [SATAC] is produced. For purposes herein, it is understood that SATACs may not necessarily be produced in multiple steps, but may appear after

20 the initial introduction of the heterologous DNA and growth under selective conditions, or they may appear after several cycles of growth under selective conditions and BrdU treatment.

As used herein an amplicon is the smallest repeated unit that contains heterologous DNA in the MACs provided herein. A megareplicon

25 contains at least one amplicon, an inverted repeat thereof, and a megareplicator. A megareplicator is a primary replication initiation site.

As used herein, the minichromosome refers to a chromosome derived from a dicentric chromosome [see, e.g., FIG. 1] that ~~is~~ contains more euchromatic than heterochromatic DNA.

As used herein, a megachromosome refers to a chromosome that, except for introduced heterologous DNA is substantially composed of heterochromatin. Megachromosomes are made of a tandem ^{array} ~~array~~ of amplicons that ^{contain} ~~contain~~ two inverted megareplicons bordered by introduced heterologous DNA [see, e.g., Figure 3 for a schematic drawing of a megachromosome]. For purposes herein, a megachromosome is about 50 to 400 Mb, generally about 250-400 Mb. Shorter variants, are also referred to ^{as a} ~~to a~~ truncated megachromosomes [about 90 to 120 or 150 Mb], dwarf megachromosomes [~ 150-200 Mb] and cell lines, and a micro-megachromosome [~ 60-90 Mb]. For purposes herein, the term megachromosome refers to the overall repeated structure based on a tandem array of repeated chromosomal segments [amplicons] that contain two inverted megareplicons bordered by any inserted heterologous DNA. The size will be specified.

As used herein, genetic therapy involves the transfer of heterologous DNA to the certain cells, target cells, of an individual afflicted with a disorder for which such therapy is sought. The DNA is introduced into the selected target cells in a manner such that the heterologous DNA is expressed and a product encoded thereby is produced. Alternatively, the heterologous DNA may in some manner mediate expression of DNA that encodes the therapeutic product, it may encode a product, such as a peptide or RNA that in some manner mediates, directly or indirectly, expression of a therapeutic product. Genetic therapy may also be used to introduce therapeutic compounds, such as TNF, that are not normally produced in the host or that are not produced in therapeutically effective amounts or at a therapeutically useful time. The heterologous DNA encoding the therapeutic product may be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof.

As used herein, heterologous or foreign DNA and RNA are used interchangeably and refer to DNA or RNA that does not occur naturally as part of the genome in which it is present or which is found in a location or locations in the genome that differ from that in which it occurs in nature. It is DNA or RNA that is not endogenous to the cell and has been exogenously introduced into the cell. Examples of heterologous DNA include, but are not limited to, DNA that encodes a gene or gene(s) of interest, introduced for purposes of gene therapy or for production of the encoded protein. Other examples of heterologous DNA include, but are not limited to, DNA that encodes traceable marker proteins, such as a protein that confers drug resistance, DNA that encodes therapeutically effective substances, such as anti-cancer agents, enzymes and hormones, and DNA that encodes other types of proteins, such as antibodies. Antibodies that are encoded by heterologous DNA may be secreted or expressed on the surface of the cell in which the heterologous DNA has been introduced.

As used herein, a therapeutically effective product is a product that is encoded by heterologous DNA that, upon introduction of the DNA into a host, a product is expressed that effectively ameliorates or eliminates the symptoms, manifestations of an inherited or acquired disease or that cures said disease.

As used herein, transgenic plants refer to plants in which heterologous or foreign DNA is expressed or in which the expression of a gene naturally present in the plant has been altered.

As used herein, operative linkage of heterologous DNA to regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences refers to the relationship between such DNA and such sequences of nucleotides. For example, operative linkage of heterologous DNA to a promoter refers to the physical relationship between the DNA

As used herein, isolated, substantially pure DNA refers to DNA fragments purified according to standard techniques employed by those skilled in the art, such as that found in Maniatis et al. [(1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY].

As used herein, vector or plasmid refers to discrete elements that are used to introduce heterologous DNA into cells for either expression of the heterologous DNA or for replication of the cloned heterologous DNA. Selection and use of such vectors and plasmids are well within the level of skill of the art.

30 either as an extrachromosomal element or by chromosomal integrant.

As used herein, injected refers to the microinjection [use of a small syringe] of DNA into a cell.

As used herein, substantially homologous DNA refers to DNA that includes a sequence of nucleotides that is sufficiently similar to another
5 such sequence to form stable hybrids under specified conditions.

It is well known to those of skill in this art, that nucleic acid fragments with different sequences may, under the same conditions, hybridize detectably to the same "target" nucleic acid. Two nucleic acid fragments hybridize detectably, under stringent conditions over a
10 sufficiently long hybridization period, because one fragment contains a segment of at least about 14 nucleotides in a sequence which is complementary [or nearly complementary] to the sequence of at least one segment in the other nucleic acid fragment. If the time during which hybridization is allowed to occur is held constant, at a value during which,
15 under preselected stringency conditions, two nucleic acid fragments with exactly complementary base-pairing segments hybridize detectably to each other, departures from exact complementarity can be introduced into the base-pairing segments, and base-pairing will nonetheless occur to an extent sufficient to make hybridization detectable. As the departure from
20 complementarity between the base-pairing segments of two nucleic acids becomes larger, and as conditions of the hybridization become more stringent, the probability decreases that the two segments will hybridize detectably to each other.

Two single-stranded nucleic acid segments have "substantially the same sequence," within the meaning of the present specification, if
25 (a) both form a base-paired duplex with the same segment, and (b) the melting temperatures of said two duplexes in a solution of 0.5 X SSPE differ by less than 10°C. If the segments being compared have the same number of bases, then to have "substantially the same sequence", they
30 will typically differ in their sequences at fewer than 1 base in 10.

Methods for determining melting temperatures of nucleic acid duplexes are well known [see, e.g., Meinkoth and Wahl (1984) Anal. Biochem. 138:267-284 and references cited therein].

- As used herein, a nucleic acid probe is a DNA or RNA fragment
- 5 that includes a sufficient number of nucleotides to specifically hybridize to DNA or RNA that includes identical or closely related sequences of nucleotides. A probe may contain any number of nucleotides, from as few as about 10 and as many as hundreds of thousands of nucleotides. The conditions and protocols for such hybridization reactions are well
- 10 known to those of skill in the art as are the effects of probe size, temperature, degree of mismatch, salt concentration and other parameters on the hybridization reaction. For example, the lower the temperature and higher the salt concentration at which the hybridization reaction is carried out, the greater the degree of mismatch that may be present in the
- 15 hybrid molecules.

- To be used as an hybridization probe, the nucleic acid is generally rendered detectable by labelling it, with a detectable moiety or label, such as ^{32}P , ^3H and ^{14}C , or by other means, including chemical labelling, such as by nick-translation in the presence of deoxyuridylate biotinylated at the
- 20 5'-position of the uracil moiety. The resulting probe includes the biotinylated uridylate in place of thymidylate residues and can be detected [via the biotin moieties] by any of a number of commercially available detection systems based on binding of streptavidin to the biotin. Such commercially available detection systems can be obtained, for example,
- 25 from Enzo Biochemicals, Inc. [New York, NY]. Any other label known to those of skill in the art, including non-radioactive labels, may be used as long as it renders the probes sufficiently detectable, which is a function of the sensitivity of the assay, the time available [for culturing cells, extracting DNA, and hybridization assays], the quantity of DNA or RNA

available as a source of the probe, the particular label and the means used to detect the label.

Once sequences with a sufficiently high degree of homology to the probe are identified, they can readily be isolated by standard techniques, which are described, for example, by Maniatis *et al.* ((1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

As used herein, conditions under which DNA molecules form stable hybrids and are considered substantially homologous are such that the DNA molecules with at least about 60% complementarity form stable hybrids. Such DNA fragments are herein considered to be "substantially homologous". For example, DNA that encodes a particular protein is substantially homologous to an other DNA fragment if the DNA forms stable hybrids such that the sequences of the fragments are at least about 60% complementary and if a protein encoded by the DNA retains its activity.

For purposes herein, the following stringency conditions are defined:

- 1) high stringency: 0.1 x SSPE, 0.1% SDS, 65°C
- 2) medium stringency: 0.2 x SSPE, 0.1% SDS, 50°C
- 3) low stringency: 1.0 x SSPE, 0.1% SDS, 50°C

or any combination of salt and temperature and other reagents that result in selection of the same degree of mismatch or matching.

As used herein, immunoprotective refers to the ability of a vaccine or exposure to an antigen or immunity-inducing agent to confer upon a host to whom the vaccine or antigen is administered or introduced the ability to resist infection by a disease causing pathogen or to have reduced symptoms. The selected antigen is typically an antigen that is presented by the pathogen.

As used herein, all assays and procedures, such as hybridization reactions and antibody-antigen reactions, unless otherwise specified, are conducted under conditions recognized by those of skill in the art as standard conditions.

5 **A. Preparation of cell lines containing MACs**

10 The methods, cells and MACs provided herein are produced by virtue of the discovery of the existence of a higher-order replication unit [megareplicon] of the centromeric region. This megareplicon is delimited by a primary replication initiation site [megareplicator], and appears to facilitate replication of the centromeric heterochromatin, and most likely, centromeres. Integration of heterologous DNA into the megareplicator region or in close proximity thereto, initiates a large-scale amplification of megabase-size chromosomal segments, which leads to *de novo* chromosome formation.

15 Cell lines containing MACs can be prepared using cells, preferably a stable cell line, transforming it with a heterologous DNA fragment that encodes a selectable marker, culturing under selective conditions, and identifying cells that have a dicentric chromosome. These cells can then be manipulated as described herein to produce the minichromosomes and other MACs, particularly the heterochromatic SATACs as described herein.

25 Development of a dicentric chromosome appears to require integration of the heterologous DNA in the pericentric heterochromatin. Thus, the probability of incorporation can be increased by including DNA, such as satellite DNA, in the heterologous fragment that encodes the selectable marker. The resulting cell lines can then be treated as the exemplified cells herein to produce cell in which the dicentric chromosome has fragmented and to introduce additional selective markers into the dicentric chromosome, whereby amplification of the pericentric heterochromatin will produce the heterochromatic

30

EC3/7
EC7/8
a chromosomes. The following discussion is with reference to the line and use of resulting cells. The same procedures can be applied and to any other cells, particularly cell lines to prepare create SATACs and euchromatic minichromosomes.

5 1. Formation of *de novo* chromosomes

De novo centromere formation in a transformed mouse LMTK-fibroblast cell line [EC3/7] after cointegration of λ constructs [λ CM8 and λ gtWESneo] carrying human and bacterial DNA [Hadlaczky et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:8106-8110 and U.S. application Serial No. 08/375,271] has been shown. The integration of the "heterologous" human and phage DNA, and the subsequent amplification of mouse and heterologous DNA that led to the formation of a dicentric chromosome, occurred at the centromeric region of the short arm of a mouse chromosome. By G-banding this chromosome was identified as mouse chromosome 7. Because of the presence of two functionally active centromeres on the same chromosome, regular breakages occur between the centromeres. Such specific chromosome breakages gave rise to the appearance [in approximately 10% of the cells] of a chromosome fragment carrying the neo-centromere. From the EC3/7 cell line [see, U.S. Patent No. 5,288,625, deposited at the European Collection of Animal cell Culture (hereinafter ECACC) under accession no. 90051001]; see, also Hadlaczky et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:8106-8110, and U.S. application Serial No. 08/375,271 and the corresponding published European application EP 0 473 253] carrying either the dicentric chromosome or a chromosome fragment with the neo-centromere, two sublines [EC3/7C5 and EC3/7C6] were selected by repeated single-cell cloning. In these cell lines, the neo-centromere was found exclusively on a minichromosome [neo-minichromosome]), while the formerly dicentric chromosome carried traces of "heterologous" DNA.

It has now been discovered that integration of DNA encoding a selectable marker in the heterochromatic region of the centromere led to formation of the dicentric chromosome.

2. The neo-minichromosome

- 5 The chromosome breakage in the EC3/7 cells, which separates the neo-centromere from the mouse chromosome, occurred in the G-band positive "heterologous" DNA region. This is supported by the observation of traces of λ and human DNA sequences at the broken end of the formerly dicentric chromosome. Comparing the G-band pattern of the chromosome
- 10 fragment carrying the neo-centromere with that of the stable neo-minichromosome, it is apparent that the neo-minichromosome is an inverted duplicate of the chromosome fragment that bears the neo-centromere. This is supported by the observation that although the neo-minichromosome carries only one functional centromere, both ends of the minichromosome
- 15 are heterochromatic, and mouse satellite DNA sequences were found in these heterochromatic regions by *in situ* hybridization.

- Mouse cells containing the minichromosome, which is composed of multiple repeats of the heterologous DNA, which in the exemplified embodiment is lambda DNA and neo DNA, can be used as recipient cells in
- 20 cell transformation. Donor DNA, such as selected heterologous DNA linked to a second selectable marker, such as hygromycin resistance [hyg], can be introduced into the mouse cells and integrated into the minichromosomes by homologous recombination of lambda DNA in the donor DNA with that in the minichromosomes. Integration is verified by *in situ* hybridization and
- 25 Southern blot analyses. Transcription and translation of the heterologous DNA is confirmed by primer extension and immunoblot analyses.

- For example, DNA has been targeted into the λ -neo minichromosome in EC3/7C5 cells using a lambda DNA-containing construct [pNem1ruc] that also contains DNA encoding hygromycin resistance and the Renilla luciferase
- 30 gene linked to a promoter, such as the cytomegalovirus [CMV] early

promoter, and the bacterial neo encoding DNA. Integration of the donor DNA into the chromosome in selected cells [designated PHN4] was confirmed by nucleic acid amplification [PCR] and *in situ* hybridization. Events that would produce a neo-minichromosome are depicted in Figure 1.

5 The resulting engineered minichromosome that contains the heterologous DNA can then be transferred by cell fusion into a recipient cell line, such as Chinese hamster kidney cells [CHO] and correct expression of the heterologous DNA can be verified. Following production of the cells, metaphase chromosomes are obtained, such as by addition of colchicine, and the chromosomes purified by addition of AT and GC specific dyes on a dual laser beam based cell sorter. Preparative amounts of chromosomes [2-3 mls of 10^6 ^{Chromosomes/ml} ~~cells/ml~~] at a purity of 95% or higher can be obtained. The resulting chromosomes are used for delivery to cells by methods, such as microinjection, liposome packaged transfer.

15 Thus, the neo-minichromosome is stably maintained in cells, replicates autonomously, and permits the persistent long-term expression of *neo* gene under non-selective culture conditions. It also contains megabases of heterologous known DNA [lambda DNA in the exemplified embodiments] that serves as target sites for homologous recombination and integration of DNA of interest. The neo-minichromosome is, thus, a vector for genetic engineering of cells.

25 The methods herein provide means to induce the events that led to formation of the neo-minichromosome by introducing heterologous DNA with a selective marker [preferably a dominant selectable marker] and culturing under selective conditions. As a result, cells that contain a dicentric chromosome or fragments thereof produced by amplification, will be produced. Cells with the dicentric chromosome can then be treated to destabilize the genome with agents, such as BrdU and/or culturing under selective conditions, resulting in cells in which the dicentric chromosome has formed two chromosomes, a so-called minichromosome, and a formerly

30

dicentric chromosome that has typically undergone amplification in the heterochromatin where the heterologous DNA has integrated to produce a generally a SATAC or a sausage chromosome [discussed below]. These cells can be fused with other cells to separate the minichromosome from the
5 formerly dicentric chromosome into different cells so that each type of MAC can be manipulated separately.

3. Preparation of SATACs

An Exemplary protocol for preparation of SATACs is illustrated in Figure 2 [particularly C, D and F] and FIGURE 4 [see, also the EXAMPLES,
10 particularly EXAMPLES 4-7].

To prepare a SATAC, the starting materials are a cell, preferably a stable cell line, such as a fibroblast cell line, and a DNA fragment that includes DNA that encodes a selective marker. To insure integration of the DNA fragment in the heterochromatin, it is preferable to start with DNA that
15 will be targeted to the pericentric heterochromatic region of the chromosome, such as λ CM8 and vectors provided herein, such as pTEMPUD [Figure 5] that include satellite DNA. After introduction of the DNA, the cells are grown under selective conditions. The resulting cells are examined and any that have dicentric chromosomes [or heterochromatic chromosomes
20 or sausage chromosomes or other such structure [see, Figure 2C, 2D, 2E and 2F] are selected.

In particular, if a cell with a dicentric chromosome is selected, it can be grown under selective conditions, or, preferably, additional DNA encoding a second selectable marker is introduced, and the cells grown under
25 conditions selective for the second marker. The resulting cells should include chromosomes that have structures similar to those depicted in Figures 2D, 2E, 2F. Cells with a structure, such as the sausage chromosome, Figure 2D, can be selected and fused with a second cell line to eliminate other chromosomes that are not of interest. If desired cells with
30 other chromosomes can be selected and treated as described herein. If a

cell with a sausage chromosome is selected, it is treated with an agent, such as BrdU, that destabilizes the chromosome so that the heterochromatic arm forms a chromosome that is substantially heterochromatin [megachromosome, see, Figure 2F]. Structures such as the gigachromosome in which the heterochromatic arm has amplified but not broken off from the euchromatic arm, will also be observed. The megachromosome is a stable chromosome. Further manipulation, such as fusions and growth in selective conditions and/or BrdU treatment or other such treatment, can lead to fragmentation of the megachromosome to form smaller chromosomes that have the amplicon as the basic repeating unit.

The megachromosome can be further fragmented in vivo using a chromosome fragmentation vector, such as pTEMPUD [see, Figure 5 and EXAMPLE 12] to ultimately produce a chromosome that comprises the smallest stable replicable unit, about 15 Mb-50 Mb, containing two to four megareplicons.

Thus, The stable chromosomes formed *de novo* that originate from the short arm of mouse chromosome 7 have been analyzed. This chromosome region shows a capacity for amplification of large chromosome segments, and promotes *de novo* chromosome formation. Large-scale amplification at the same chromosome region leads to the formation of dicentric and multicentric chromosomes, a minichromosome, the 150-200 Mb size λ neo-chromosome, the "sausage" chromosome, the 500-1000 Mb gigachromosome, and the stable 250-400 Mb megachromosome.

A clear segmentation is observed along the arms of the megachromosome, and analyses show that the building units of this chromosome are amplicons of ~30 Mb composed of mouse major satellite DNA with the integrated "foreign" DNA sequences at both ends. The ~30 Mb amplicons are composed of two ~15 Mb inverted doublets of ~7.5 Mb mouse major satellite DNA blocks, which are separated from each other by a narrow band of non-satellite sequences [see, e.g., Figure 3]. The wider

non-satellite regions at the amplicon borders contain integrated, exogenous [heterologous] DNA, while the narrow bands of non-satellite DNA sequences within the amplicons are integral parts of the pericentric heterochromatin of mouse chromosomes. These results indicate that the ~ 7.5 Mb blocks
 5 flanked by non-satellite DNA are the building units of the pericentric heterochromatin of mouse chromosomes, and the ~ 15 Mb size pericentric regions of mouse chromosomes contain two ~ 7.5 Mb units.

Apart from the euchromatic terminal segments, the whole megachromosome is heterochromatic, and has structural homogeneity.
 10 Therefore, this large chromosome offers a unique possibility for obtaining information about the amplification process, and for analyzing some basic characteristics of the pericentric constitutive heterochromatin, as vector for heterologous DNA, and as target for further fragmentation.

As shown herein, this phenomenon is generalizable and can be
 15 observed with other chromosomes. Also, although these *de novo* formed chromosome segments and chromosomes appear different, there are, similarities that indicate that a similar amplification mechanism plays a role in their formation: (i) in each case, the amplification is initiated in the centromeric region of the mouse chromosomes and large (Mb size)
 20 amplicons are formed; (ii) mouse major satellite DNA sequences are constant constituents of the amplicons, either by providing the bulk of the heterochromatic amplicons [H-type amplification], or by bordering the euchromatic amplicons [E-type amplification]; (iii) formation of inverted segments can be demonstrated in the λ neo-chromosome and
 25 megachromosome; (iv) chromosome arms and chromosomes formed by the amplification are stable and functional.

The presence of inverted chromosome segments seems to be a common phenomenon in the chromosomes formed *de novo* at the centromeric region of mouse chromosome 7. During the formation of the
 30 neo-minichromosome, the event leading to the stabilization of the distal

segment of mouse chromosome 7 that bears the neo-centromere may have been the formation of its inverted duplicate. Amplicons of the megachromosome are inverted doublets of ~7.5 Mb mouse major satellite DNA blocks.

5 4. Cell lines

Cell lines that contain MACs, such as the minichromosome, the λ -neo chromosome, and the SATACs are provided herein or can be produced by the methods herein. Such cell lines provide a convenient source of these chromosomes and can be manipulated, such as by cell fusion or production of microcells for fusion with selected cell lines, to deliver the chromosome of interest into hybrid cell lines. Exemplary cell lines are described herein and some have been deposited with the ECACC.

a. EC3/7C5 and EC3/7C6

Two cell lines EC3/7C5 and EC3/7C6 produced by single cell cloning of EC3/7 were examined. For exemplary purposes EC3/7C5 has been deposited with the ECACC. These cell lines contain a minichromosome and the formerly dicentric chromosome from EC3/7. The stable minichromosomes in cell lines [EC3/7C5 and EC3/7C6] appear to be virtually identical and they seem to be a duplicated derivatives of the ~10-15 Mb "broken-off" fragment of the dicentric chromosome. Their identical size in these independently generated cell lines might indicate that ~20-30 Mb is the minimal or close to the minimal physical size for a stable minichromosome.

b. TF1004G/19

Introduction of additional heterologous DNA, including a second selectable marker hygromycin and also a detectable marker β -galactosidase, into the EC3/7C5 cell line and growth under selective conditions produced cells designated TF1004G/19. In particular, this cell line was produced from the EC3/7C5 cell line by cotransfection with plasmids pH132, which contains an anti-HIV ribozyme, and hygromycin resistance gene, pCH110

[encodes β -galactosidase] and λ phage [λ cl 875 Sam 7] DNA and selection with hygromycin B.

Detailed analysis of TF1004G/19 cell line by *in situ* hybridization with lambda phage and plasmid DNA sequences revealed the formation of the
 5 sausage chromosome. The formerly dicentric chromosome of EC3/7C5 cell translocated to the end of another acrocentric chromosome. The heterologous DNA integrated into the pericentric heterochromatin of formerly dicentric chromosome and is amplified several times with megabases of
 10 forming the "sausage" chromosome. Subsequently the acrocentric mouse chromosome was substituted by a euchromatic telomere.

In situ hybridization with biotin labeled subfragments of the hygromycin resistance and β -galactosidase genes resulted in hybridization
 15 signal only in the heterochromatic arm of the sausage chromosome, indicating that in TF1004G/19 transformant cells these genes are localized in the pericentric heterochromatin. A high level of gene expression, however, was detected.

In general, heterochromatin has a silencing effect in *Drosophila*, yeast and on the HSV-tk gene introduced into satellite DNA at mouse centromere.
 20 Thus, it was of interest to study the TF1004G/19 transformant cell line in to confirm that gene expression [of the β -gal] was indeed localized in the heterochromatin contrary to recognized dogma.

For this purpose, subclones of TF1004G/19 containing a different a sausage chromosome [see Figure 2D] by single cell cloning were estab-
 25 lished. Southern DNA hybridization with subfragments of hygromycin resistance and β -galactosidase genes showed close correlation to the intensity of hybridization and the length of sausage chromosome. This finding supports the conclusion that these genes are localized in the heterochromatic arm of the sausage chromosome.

TF1004G-19C5 is a mouse LMTK⁻ fibroblast cell line containing neo-minichromosomes and stable "sausage" chromosomes. It is a subclone of TF1004G/19. It has been deposited as an exemplary cell line and exemplary source of a sausage chromosome. Subsequent fusion of this cell line with CHO K20 cells and selection with hygromycin and hat resulted in hybrid cells that carry the sausage chromosome and/or the neo-minichromosome. BrdU treatment, single cell cloning and selection with G418 and/or hygromycin produced various cells that carry chromosomes of interest, including G3D5.

10 G3D5, which has been deposited is a mouse hamster hybrid cell line that carries the neo-minichromosome and the megachromosome. H1D3 is a subclone thereof that carries the megachromosome. Fusion of this cell line with the CD4⁺ Hela cell line and also DNA encoding an additional selection
15 gene, neo, produced cells that carry the megachromosome as well as a human chromosome that carries CD4neo [H1D3 cells]. Further BrdU treatment and single cell cloning produced cell lines, such as 1B3 that include cells with a truncated megachromosome.

5. DNA constructs used to transform cells by other suitable method at any stage during preparation of the chromosomes [see, e.g., FIG. 4]. In general integration of such DNA is assured by relying on site directed integration, such as by inclusion of λ -DNA for the exemplified chromosomes and also an additional selective marker gene. For example, cells with a MAC, such as the minichromosome or a SATAC can be cotransfected with a plasmid encoding the desired heterologous DNA, such as HIV ribozyme, cystic fibrosis gene, and a second selectable marker, such as hygromycin resistance. Selection is effected with the agent that selects for the new selectable marker cells containing chromosomes that include the DNA in the MAC are identified. Fusion with a second cell line

can provide a means to produce cell lines that contain one particular type of chromosomal structure or MAC.

Various vectors for this purpose are provided herein [see, Examples] and others can be readily constructed. The vectors should include DNA that will target the DNA to the MAC, a selectable marker and the selected heterologous gene of interest. Based on the disclosure herein and the knowledge of the skilled artisan, one of skill can construct such vectors.

Of particular interest herein is the vector pTEMPUD and derivatives thereof that can target DNA into the heterochromatic region of selected chromosomes. These vectors can also serve as fragmentation vectors [see, e.g., Example 12].

Heterologous genes of interest include any gene that encodes a therapeutic gene and DNA encoding gene products of interest. These genes and DNA include, but are not limited to: the cystic fibrosis gene [CF] cystic fibrosis transmembrane regulator (CFTR) [see, e.g., U.S. Patent No. 5,240,846; Rosenfeld *et al.* (1992) *Cell* 68:143-155; Hyde *et al.* (1993) *Nature* 362: 250-255; Kerem *et al.* (1989) *Science* 245:1073-1080; Riordan *et al.* (1989) *Science* 245:1066-1072; Rommens *et al.* (1989) *Science* 245:1059-1065; Osborne *et al.* (1991) *Am. J. Hum. Genetics* 48:6089-6122; White *et al.* (1990) *Nature* 344:665-667; Dean *et al.* (1990) *Cell* 61:863-870; Erlich *et al.* (1991) *Science* 252:1643; and U.S. Patent Nos. 5,453,357, 5,449,604, 5,434,086, and 5,240,846, which provides a retroviral vector encoding the normal CFTR gene].

B. Isolation of artificial chromosomes

The MACs provided herein can be isolated by any suitable method known to those of skill in the art. Also, a method is provided herein for effecting substantial purification. SATACs have been isolated by fluorescence activated cell sorting [FACS]. This method takes advantage of the nucleotide base content of the SATACs, which by virtue of their heterochromatic content will differ from any other chromosomes in a cell.

In particular, metaphase chromosomes are isolated and stained with base specific dyes, such as Hoechst 33258 and chromocycin A3. Fluorescence activated cell sorting will separate the SATACs from the genomic chromosomes. A dual-laser cell sorter [FACStar Plus and FAXStar Vantage
 5 Becton Dickinson Immunocytometry System] in which two lasers were set to excite the dyes separately, allowed a bivariate analysis of the chromosomes by size and base-pair composition. Cells containing such SATACs can be similarly sorted.

10 C. Introduction of artificial chromosomes into cells, tissues, animals and plants

Suitable hosts for introduction of the MACs provided herein, include but are not limited to any animal or plant, cell or tissue thereof, including, but not limited to: mammals, birds, reptiles, amphibians, insects, fish, arachnids, tobacco, tomato, wheat, monocots, dicots and algae. The MACs may be
 15 introduced by cell fusion or microcell fusion or subsequent to isolation by any method known to those of skill in this art, including but not limited to: direct DNA transfer, electroporation, lipofection, liposomes, microprojectile bombardment, microinjection and any other suitable method.

Other methods for introducing DNA into cells, include nuclear
 20 microinjection, electroporation, bacterial protoplast fusion with intact cells. Polycations, such as polybrene and polyornithine, may also be used. For various techniques for transforming mammalian cells, see e.g., Keown et al. Methods in Enzymology (1990) Vol. 185, pp. 527-537; and Mansour et al. (1988) Nature 336:348-352.

25 DNA may be introduced by direct DNA transformation; microinjection in cells or embryos, protoplast regeneration for plants, electroporation, microprojectile gun and other such methods [see, e.g., Weissbach et al. (1988) Methods for Plant Molecular Biology, Academic Press, N.Y., Section VIII, pp. 421-463; Grierson et al. (1988) Plant Molecular Biology, 2d Ed.,
 30 Blackie, London, Ch. 7-9; see, also U.S. Patent Nos. 5,491,075; 5,482,928; and 5,424,409; see, also, e.g., U.S. Patent No. 5,470,708,

cells].

For example, isolated purified artificial chromosomes can be injected into an embryonic cell line such as a human kidney primary embryonic cell line [ATCC CRL 1573] or embryonic stem cells [see, e.g., Hogan et al. (1994) *Manipulating the Mouse Embryo, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, see, especially, pages 255-264 and Appendix 3]. Preferably the chromosomes are introduced by microinjection, using a system such as the Eppendorf automated microinjection system, and grown under selective conditions, such as hygromycin B or neomycin resistance.

15 Depending on the host cell used, transformation is done using standard techniques appropriate to such cells. These methods include any, including those described herein, known to those of skill in the art.

For mammalian cells without such cell walls, the calcium phosphate precipitation method [see, e.g., Graham et al. (1978) *Virology* 52:456-457] is often preferred. DNA uptake can be accomplished by DNA alone or in the presence of polyethylene glycol [PEG-mediated gene transfer], which is a fusion agent, with plant protoplasts or by any variations of such methods known to those of skill in the art [see, et al. U.S. Pat. No. 4,684,611].

-32-

b. Electroporation

Electroporation, which involves providing high-voltage electrical pulses to a solution containing a mixture of protoplasts and foreign DNA to create reversible pores in the membranes of plant protoplasts as well as other cells. Electroporation is generally used for prokaryotes or other cells, such as plants that contain substantial cell-wall barriers. Methods for effecting electroporation are well known [see, e.g., U.S. Patent Nos. 4,784,737, 5,501,967, 5,501,662, 5,019,034, 5,503,999; see, also Frommet et al. (1985) Proc. Natl. Acad. Sci. U.S.A. 82:5824-5828].

For example, electroporation is often used for transformation of plants [see, e.g., Ag Biotechnology News, Vol. 7 p. 3 and 17 (September/October 1990)]. In this technique, plant protoplasts are electroporated in the presence of the DNA of interest that also includes a phenotypic marker. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus. Transformed plant cells will be identified by virtue of the expressed phenotypic marker. The exogenous DNA may be added to the protoplasts in any form such as, for example, naked linear, circular or supercoiled DNA, DNA encapsulated in liposomes, DNA in spheroplasts, DNA in other plant protoplasts, DNA complexed with salts, and other methods.

c. Microcells

The chromosomes can be transferred by preparing microcells containing an artificial chromosome and then fusing with selected target cells. Methods for such preparation and fusion or microcells are well known [see, e.g., U.S. Patent Nos. 5,240,840, 4,806,476, 5,298,429, Fournier (1981) Proc. Natl. Acad. Sci. U.S.A. 78:6349-6353; and Lambert et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:5907-59].

2. Hosts

Suitable host include any host known to be useful for introduction and expression of heterologous DNA. Of particular interest herein, animal and plant cells and tissues, including, but not limited to insect cells and larvae, plants, and animals, particularly transgenic animals, and animal cells. Other hosts include, but are not limited to mammals, birds, reptiles, amphibians, insects, fish, arachnids, tobacco, tomato, wheat, monocots, dicots and algae, and any host into which introduction of heterologous DNA is desired.

Such introduction can be effected using the MACs provided herein, or, if necessary by using the MACs provided herein to identify species-specific centromeres and/or functional chromosomal units and then using the resulting centromeres or chromosomal units as artificial chromosomes, or alternatively, using the methods exemplified herein for production of MACs to produce species-specific artificial chromosomes.

15 a. **Introduction of DNA into embryos for production of transgenic animals and introduction of DNA into animal cells**

Transgenic animals can be produced by introducing exogenous genetic material into a pronucleus of a mammalian zygote by microinjection [see, e.g., U.S. Patent Nos. 4,873,191 and 5,354,674; see, also, International PCT application No. WO95/14769, which is based on U.S. application Serial No. 08/159,084]. The zygote is capable of development into a mammal. The embryo or zygote is transplanted into a host female uterus and allowed to develop. Detailed protocols and examples are set forth below.

DNA can be introduced into animal cells using any known procedure, including, but not limited to: direct uptake, incubation with polyethylene glycol [PEG], microinjection, electroporation, lipofection, cell fusion, microcell fusion, particle bombardment, including microprojectile bombardment [see, e.g., U.S. Patent No. 5,470,708, which provides a method for transforming unattached mammalian cells via particle bombardment], and any other such method. For example, the transfer of plasmid DNA in liposomes directly to

b. Introduction of heterologous DNA into plants.

- Plant species, including tobacco, rice, maize, rye, soybean, Brassica napus, cotton, lettuce, potato and tomato, have been used to produce transgenic plants. Tobacco and other species, such as petunias, often serve as experimental models in which the methods have been developed and the genes first introduced and expressed.

- 35-

All plant cells into which DNA can be introduced and that can be regenerated from the transformed cells can be used to produce transformed whole plants which contain the transferred chromosome. The particular protocol and means for introduction of the DNA into the plant host may need
 5 to be adapted or refined to suit the particular plant species or cultivar. c .

Insect cells

Insects are useful hosts for introduction of artificial chromosomes for numerous reasons, including, but not limited to: (a) amplification of genes encoding useful proteins can be accomplished in the artificial mammalian
 10 chromosome to obtain higher protein yields in insect cells; (b) insect cells support required post translational modifications, such as glycosylation and phosphorylation, that can be required for protein biological functioning; (c) insect cells do not support mammalian viruses, and, thus, eliminate the problem of cross-contamination of products with such infectious agents; (d)
 15 this technology circumvents traditional recombinant baculovirus systems for production of nutritional, industrial or medicinal proteins in insect cell systems; (e) the low temperature optimum for insect cell growth (28° C) permits reduced energy cost of production; (f) serum free growth medium for insect cells permits lower production costs; (g) artificial chromosome-
 20 containing cells can be stored indefinitely at low temperature; and (h) insect larvae will be biological factories for production of nutritional, medicinal or industrial proteins by microinjection of fertilized insect eggs [see, e.g., Joy et al. (1991) Current Science 66:145-150, which provides a method for microinjecting heterologous DNA into *Bombyx mori* eggs].

25 Either MACs or insect-specific [BUGACs] will used to introduce genes into insects. As described in the Examples, it appears that MACs will function in insects to direct expression of heterologous DNA contained thereon.

Insect host cells include, but are not limited to, hosts such as
 30 *Spodoptera frugiperda* [caterpillar], *Aedes aegypti* [mosquito], *Aedes*

- albopictus* [mosquito], *Drosophila melanogaster* [fruitfly], *Bombyx mori* [silkworm] and *Trichoplusia ni* [cabbage looper]. Effort has been directed toward propagation of insect cells in culture. Such efforts have focused on the fall armyworm, *Spodoptera frugiperda*. The cell lines have been
- 5 developed also from other insects such as the cabbage looper, *trichoplusia ni* and the silkworm, *Bombyx mori*. It has also been suggested that analogous cell lines can be created using the tobacco hornworm, or *Manduca sexta*. To introduce DNA into an insect, it should be introduced into the larvae, and allowed to proliferate, and then the hemolymph
- 10 recovered from the larvae so that the proteins can be isolated therefrom. The preferred method herein is microinjection [see, e.g., Tamura et al. (1991) Bio Ind. 8:26-31; Nikolaev et al. (1989) Mol. Biol. (Moscow) 23:1177-87; and methods exemplified and discussed herein].

D. Applications for and Uses of Artificial chromosomes

- 15 Artificial chromosomes provide convenient and useful vectors, and in some instances the only vectors for introduction of heterologous genes into hosts. Virtually any gene of interest is amenable to introduction into a host via an artificial chromosome. Such genes include, but are not limited to genes that encode receptors, cytokines, enzymes, proteases, hormones,
- 20 growth factors, tumor suppressor genes, therapeutic products and multigene pathways.

- The artificial chromosomes provided herein will be used in methods of protein and gene product production, particularly using insects as host cells for production of such products. They are also intended for use in
- 25 methods of gene therapy, and in for production of transgenic plants and animals [discussed above, below and in the EXAMPLES].

1. Gene Therapy

- Any therapeutic gene product or product of a multigene pathway may be introduced into a host animal, such as a human, or into a target cell line
- 30 for introduction into an animal, for therapeutic purposes. Such therapeutic

purposes include, genetic therapy to cure or to provide gene products that are missing or defective, to deliver agents, such as anti-tumor agents, to targeted cells or to an animal, and to provide gene products that will confer resistance or reduce susceptibility to a pathogen or ameliorate symptoms of a disease or disorder. The following are some exemplary genes and gene products. Such exemplification is not intended to be limiting.

a. Anti-HIV ribozymes

As exemplified below, DNA encoding anti-HIV ribozymes can be introduced and expressed using MACs, including the euchromatin-based minichromosomes and the SATACs. These MACs can be used to make a transgenic mouse with that express a ribozyme and, thus, serve as a model for testing the activity of such ribozymes or from which ribozyme-producing cell lines can be made. Also, introduction of a MAC into human cells that encodes an anti-HIV ribozyme will serve as treatment for HIV infection.

b. Tumor Suppressor Genes

Tumor suppressor genes are genes that, in their wild-type alleles, express proteins that suppress abnormal cellular proliferation. When the gene coding for a tumor suppressor protein is mutated or deleted, the resulting mutant protein or the complete lack of tumor suppressor protein expression may fail to correctly regulate cellular proliferation, and abnormal cellular proliferation may take place, particularly if there is already existing damage to the cellular regulatory mechanism. A number of well-studied human tumors and tumor cell lines have been shown to have missing or nonfunctional tumor suppressor genes.

Examples of tumor suppression genes include, but are not limited to, the retinoblastoma susceptibility gene or RB gene, the p53 gene, the deleted in colon carcinoma [DCC] gene and the neurofibromatosis type 1 [NF-1] tumor suppressor gene [see, e.g., U.S. Patent No. 5,496,731; Weinberg *et al.* (1991) 254:1138-1146]. Loss of function or inactivation of tumor

suppressor genes may play a central role in the initiation and/or progression of a significant number of human cancers.

The p53 Gene

Somatic cell mutations of the p53 gene are said to be the most frequently mutated gene in human cancer [see, e.g., Weinberg et al. (1991) Science 254:1138-1146]. The normal or wild-type p53 gene is a negative regulator of cell growth, which, when damaged, favors cell transformation. The p53 expression product is found in the nucleus, where it may act in parallel with or cooperatively with other gene products. Tumor cell lines in which p53 has been deleted have been successfully treated with wild-type p53 vector to reduce tumorigenicity [see, Baker et al. (1990) Science 249:912-915].

DNA encoding the p53 gene and plasmids containing this DNA are well known [see, e.g., U.S. Patent No. 5,260,191; see, also Chen et al. (1990) Science 250:1576; Farrel et al. (1991) EMBO J. 10:2879-2887, plasmids containing the gene are available from the ATCC, and the sequence is in the GenBank Database, accession nos. X54156, X60020, M14695, M16494, K03199].

c. The CFTR gene

Cystic fibrosis [CF] is an autosomal recessive disease that affects epithelia of the airways, sweat glands, pancreas, and other organs. It is a lethal genetic disease associated with a defect in Cl transport, and is caused by mutations in the gene coding for cystic fibrosis transmembrane conductance regulator [CFTR], a 1480 amino acid protein that has been associated with the expression of chloride conductance in a variety of eukaryotic cell types. Defects in CFTR destroy or reduce the ability of epithelial cells in the airways, sweat glands, pancreas and other tissues to secrete Cl in response to cAMP-mediated agonists and impair activation of apical membrane channels by cAMP-dependent protein kinase A [PKA]. Given the high incidence and devastating nature of this disease, development of effective CF treatments is imperative.

The CFTR gene [~ 600 kb] can be transferred, such as from a CF-YAC [see Green et al. Science 250:94-98] by construction of a selectable CF YAC by inserting a selectable marker, such as puromycin or hygromycin resistance and λ -DNA by site specific integration into the neo-minichromosome or into a SATAC. The CF-YAC can be introduced into
 5 cells, such as EC3/7C5 or 19C5xHa4 by yeast protoplast fusion or microinjection of yeast nuclei into mammalian cells, select stable transformants, and establish antibiotic-resistant cell lines.

2. Disease resistant animals and plants

10 Artificial chromosomes are ideally suited for preparing disease resistant animals, including vertebrates and invertebrates, including fish as well as mammals. In particular, multivalent vaccines can be prepared. Such vaccines will be encoded by multiple antigens that can be included in
 15 a MAC and either delivered to a host to induce immunity, or incorporated into embryos to produce disease-resistant animals and plants [or plants and animals that are less susceptible].

Fish and crustaceans will serve as model hosts for production of disease resistant animals.

3. Use of MACs and other artificial chromosomes for preparation of and screening of libraries

20 Since large fragments can incorporated into each SATAC, entire genomes can be readily screened. For example, DNA encoding tree growth factors can be introduced into trees. Libraries can be prepared, introduce large fragments into chromosomes, and introduce them all into trees, thereby
 25 insuring expression.

The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

EXAMPLE 1

General Materials and Methods

30 The following materials and methods are exemplary of methods that are used in the following Examples and that can be used to pre. Other

suitable materials and methods known to those of skill in the art may used. Modifications of these materials methods known to those of skill in the art may also be employed.

A. Culture of cell lines, cell fusion, and transfection of cells

- 5 1. Chinese hamster K20 cells and mouse A9 fibroblast cells were cultured in F-12 medium. EC3/7 [see, U.S. Patent No. 5,288,625, and deposited at the European Collection of Animal cell Culture (ECACC) under accession no. 90051001; see, also Hadlaczky *et al.* (1991) Proc. Natl. Acad. Sci. U.S.A. 88:8106-8110 and U.S. application Serial No. 08/375,271] and
- 10 EC3/7C5 [see, U.S. Patent No. 5,288,625 and Praznovszky *et al.* (1991) Proc. Natl. Acad. Sci. U.S.A. 88:11042-11046] mouse cell lines, and the KE1-2/4 hybrid cell line were maintained in F-12 medium containing 400 μ g/ml G-418 [SIGMA, St. Louis, MO].
- 15 2. TF1004G19 and TF1004G19C5 mouse cells, described below, and the 19xHa4 hybrid, described below, and its sublines were cultured in F-12 medium containing 400 μ g/ml Hygromycin B [Calbiochem]. LP11 cells were maintained in F-12 medium containing 3-15 μ g/ml Puromycin [SIGMA, St. Louis, MO].
- 20 3. Cotransfection of EC3/7C5 cells with plasmids [pH132, pCH110 available from Pharmacia, see, also Hall *et al.* (1983) J. Mol. Appl. Gen. 2:101-109] and with λ DNA was made using the calcium phosphate DNA precipitation method [see, *e.g.*, Chen *et al.* (1987) Mol. Cell. Biol. 7:2745-2752], using 2-5 μ g plasmid DNA and 20 μ g λ phage DNA per 5×10^6 recipient cells.
- 25 4. Cell fusion
- Mouse and hamster cells were fused using polyethylene glycol [Davidson *et al.* (1976) Som. Cell Genet. 2:165-176]. Hybrid cells were selected in HAT medium containing 400 μ g/ml Hygromycin B.
- 30 Approximately 2×10^7 recipient and 2×10^6 donor cells were fused using polyethylene glycol [Davidson *et al.* (1976) Som. Cell Genet. 2:165-176].

Hybrids were selected and maintained in F-12/HAT medium [Szybalsky *et al.* (1962) *Natl. Cancer Inst. Monogr.* 7:75-89] containing 10% FCS and 400 μ g/ml G418. The presence of "parental" chromosomes in the hybrids cell lines was verified by *in situ* hybridizations with species-specific probes using

5 biotin labeled human and hamster genomic DNA, and a mouse long interspersed repetitive DNA [pMCPE1.51].

5. Microcell fusion

Microcell-mediated chromosome transfer was done according to Saxon *et al.* [(1985) *Mol. Cell. Biol.* 1:140-146] with the modifications of

10 Goodfellow *et al.* [(1989) Techniques for mammalian genome transfer. In *Genome Analysis a Practical Approach*. K.E. Davies, ed., IRL Press, Oxford, Washington DC. pp.1-17] and Yamada *et al.* [(1990) *Oncogene* 5:1141-1147]. Briefly, 5×10^6 EC3/7C5 cells in a T25 flask were treated first with 0.05 μ g/ml colcemid for 48 hr and then with 10 μ g/ml cytochalasin

15 B for 30 min. The T25 flasks were centrifuged on edge and the pelleted microcells were suspended in serum free DME medium. The microcells were filtered through first a 5 micron and then a 3 micron polycarbonate filter, treated with 50 μ g/ml of phytohemagglutinin, and used for polyethylene glycol mediated fusion with recipient cells. Selection of cells containing the

20 MMCneo was started 48 hours after fusion in medium containing 400-800 μ g/ml G418.6

B. Chromosome banding

Trypsin G-banding of chromosomes was performed using the method of Wang & Fedoroff [(1972) *Nature* 235:52-54], and the detection of

25 constitutive heterochromatin with the BSG. C-banding method was done according to Sumner [(1972) *Cell Res.* 75:304-306]. For the detection of chromosome replication by bromodeoxyuridine [BrdU] incorporation, the Fluorescein Plus Giemsa [FPG] staining method of Perry & Wolff [(1974) *Nature* 251:156-158] was used.

30 C. Immunolabelling of chromosomes and *in situ* hybridization

Indirect immunofluorescence labelling with human anti-centromere serum LU851 [Hadlaczký *et al.* (1986) *Exp. Cell Res.* **167**:1-15, and indirect immunofluorescence and *in situ* hybridization on the same preparation were performed as described previously [see, Hadlaczký *et al.* (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**:8106-8110, see, also U.S. application Serial No. 08/375,271]. Immunolabelling with fluorescein-conjugated anti-BrdU monoclonal antibody [Boehringer] was performed according to the procedure recommended by the manufacturer, except that for treatment of mouse A9 chromosomes, 2 M hydrochloric acid was used at 37° C for 25 min, and for chromosomes of hybrid cells, 1 M hydrochloric acid was used at 37° C for 30 min.

D. Scanning electron microscopy

Preparation of mitotic chromosomes for scanning electron microscopy using osmium impregnation was performed as described previously [Sumner (1991) *Chromosoma* **100**:410-418]. The chromosomes were observed with a Hitachi S-800 field emission scanning electron microscope operated with an accelerating voltage of 25 kV.

E. DNA manipulations, plasmids and probes

1. General methods

All general DNA manipulations were performed by standard procedures [see, *e.g.*, Sambrook *et al.* (1989) *Molecular cloning: A Laboratory Manual* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY]. The mouse major satellite probe was provided by Dr. J. B. Rattner [University of Calgary, Alberta, Canada]. Cloned mouse satellite DNA probes [see, Wong *et al.* (1988) *Nucl. Acids Res.* **16**:11645-11661], including the mouse major satellite probe, were gifts from Dr. J. B. Rattner, University of Calgary. Hamster chromosome painting was done with total hamster genomic DNA, and a cloned repetitive sequence specific to the centromeric region of the chromosome 2 [Fátyol *et al.* (1994) *Nucl. Acids Res.* **22**:3728-3736] was also used. Mouse chromosome painting was done

with a cloned long interspersed repetitive sequence [pMCP1.51] specific for the mouse euchromatin.

For cotransfection and for *in situ* hybridization, the pCH110 β -galactosidase construct [Pharmacia or Invitrogen], and λ cl 875 Sam7 phage

5 DNA [New England Biolabs] were used.

2. Construction of Plasmid pPuroTel

Plasmid pPuroTel, which carries a Puromycin resistance gene and a cloned 2.5 kb human telomeric sequence [see, SEQ ID No 3], was constructed from the pBabe-puro retroviral vector [Morgenstern *et al.* (1990) 10 *Nucl. Acids Res.* 18:3587-3596; provided by Dr. L. Székely (Microbiology and Tumorbiology Center, Karolinska Institutet, Stockholm); see, also Tonghua *et al.* (1995) *Chin. Med. J.* (Beijing, Engl. Ed.) 108:653-659; Couto *et al.* (1994) *Infect. Immun.* 62:2375-2378; Dunckley *et al.* (1992) *FEBS Lett.* 296:128-34; French *et al.* (1995) *Anal. Biochem.* 228:354-355; Liu *et al.* (1995) *Blood* 85:1095-1103; International PCT application Nos. WO 9520044; WO 9500178, and WO 9419456].

F. Deposited cell lines

Cell lines KE1 2/4, EC3/7C5, TF1004G19C5, 19C5xHa4, G3D5 and H1D3 and have been deposited in accord with the Budapest Treaty at the 20 European Collection of Animal cell Culture (ECACC) under Accession Nos. 96040924, 96040925, 96040926, 96040927, 96040928 and 96040929, respectively.

EXAMPLE 2

Preparation EC3/7, EC3/7C5 and related cell lines

25 The EC3/7 cell line is an LMTK⁻ mouse cell line that contains the neo-centromere. The EC3/7C5 cell line is a single-cell subclone of EC3/7 that contains the neo-minichromosome.

A. EC3/7 Cell line

- As described in U.S. Patent No. 5,288,625 [see, also Praznovszky *et al.* (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:11042-11046 and Hadlaczký *et al.* (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:8106-8110] *de novo* centromere formation occurs in a transformed mouse LMTK-fibroblast cell line [EC3/7] after cointegration of λ constructs [λ CM8 and λ gtWESneo] carrying human and bacterial DNA.

- By cotransfection of a 14 kb human DNA fragment cloned in lambda [λ CM8] and a dominant marker gene [λ gtWESneo], a selectable centromere linked to a dominant marker gene [neo-centromere] was formed in mouse LMTK cell line EC3/7 [Hadlaczký *et al.* (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:8106-8110, see Figure 1] Integration of the heterologous DNA [the λ DNA and marker gene-encoding DNA] occurred into the short arm of an acrocentric chromosome [chromosome 7 (see, Figure 1B)], where an amplification process resulted in the formation of the new centromere [neo-centromere (see Figure 1C)]. On the dicentric chromosome (Figure 1C) the newly formed centromere region contains all the heterologous DNA (human, lambda, and neo) introduced into the cell, and an active centromere.

- Having two functionally active centromeres on the same chromosome causes regular breakages between the centromeres [see, Figure 1E]. The distance between the two centromeres on the dicentric chromosome is estimated to be ~10-15 Mb, and the breakage that separates the minichromosome occurred between the two centromeres. Such specific chromosome breakages result in the appearance [in approximately 10% of the cells] of a chromosome fragment that carries the neo-centromere [Figure 1F]. This chromosome fragment is principally composed of human, lambda, plasmid, and neo gene DNA, but it also has some mouse chromosomal DNA. Cytological evidence suggests that during the stabilization of the MMCneo, there was an inverted duplication of the chromosome fragment bearing the

neo-centromere. The size of minichromosomes in both cell lines is approximately 20-30 Mb; this finding indicates a two-fold increase in size.

From the EC3/7 cell line, which contains the dicentric chromosome [Figure 1E], two sublines [EC3/7C5 and EC3/7C6] were selected by repeated single-cell cloning. In these cell lines, the neo-centromere was found exclusively on a small chromosome [neo-minichromosome], while the formerly dicentric chromosome carried detectable amounts of the exogenously-derived DNA sequences but not an active neo-centromere [Figure 1F and 1G].

The minichromosomes of cell lines EC3/7C5 and EC3/7C6 are similar. No differences are detected in their architectures at either the cytological or molecular level. The minichromosomes were indistinguishable by conventional restriction endonuclease mapping or by long-range mapping using pulsed field electrophoresis and Southern hybridization. The cytoskeleton of cells of the EC3/7C6 line showed an increased sensitivity to colchicine, so the EC3/7C5 line was used for further detailed analysis.

B. Preparation of the EC3/7C5 and EC3/7C6 cell lines

Integration of the "foreign" DNA, and subsequent amplification of mouse and "foreign" DNA that leads to the formation of a dicentric chromosome, occurs at the centromeric region of the short arm of a mouse chromosome, identified by G-banding as mouse chromosome 7. Because of the presence of two functionally active centromeres on the same chromosome, regular breakages occur between the centromeres. Such specific chromosome breakages give rise to the appearance [in approximately 10% of the cells] of a chromosome fragment carrying the neo-centromere and the remaining portion of the formerly dicentric chromosome. From the EC3/7 cell line carrying either the dicentric chromosome or a chromosome fragment with the neo-centromere, two sublines [EC3/7C5 and EC3/7C6] were selected by repeated single-cell cloning. In these cell lines, the neo-centromere was found exclusively on a

minichromosome (neo-minichromosome), while the formerly dicentric chromosome carried traces of "foreign" DNA sequences.

- a* The ^{EC37/C5}EC37/C5 cells, which contain the dicentric chromosome, were produced by subcloning the EC3/7 cell line in high concentrations of G418 [40-fold the lethal dose] for 350 generations. Two single cell-derived stable cell lines [EC3/7C5 and EC3/7C6] were established. These cell lines carry the ^{neo-centromere}human centromere on minichromosomes and also contain the remaining fragment of the dicentric chromosome. Indirect immunofluorescence with anti-centromere antibodies and subsequent in situ hybridization experiments demonstrated that the minichromosomes derived from the dicentric chromosome. In EC3/7C5 and EC3/7C6 cell lines (140 and 128 metaphases, respectively) no dicentric chromosomes were found and minichromosomes were detected in 97.2% and 98.1% of the cells, respectively. The minichromosomes have been maintained for over 150 cell generations. They do contain the remaining portion of the formerly dicentric chromosome.

- Multiple copies of telomeric DNA sequences were detected in the marker centromeric region of the dicentric chromosome by in situ hybridization. This indicates that mouse telomeric sequences were coamplified with the foreign DNA sequences. These stable minichromosome-carrying cell lines provide direct evidence that the extra centromere containing human DNA is functioning and is capable of maintaining the minichromosomes [see, U.S. Patent No. 5,288,625].

- The chromosome breakage in the EC3/7 cells, which separates the neo-centromere from the mouse chromosome, occurred in the G-band positive "foreign" DNA region. This is supported by the observation of traces of λ and human DNA sequences at the broken end of the formerly dicentric chromosome. Comparing the G-band pattern of the chromosome fragment carrying the neo-centromere with that of the stable neo-minichromosome, reveals that the neo-minichromosome is an inverted

duplicate of the chromosome fragment that bears the neo-centromere. This is also evidenced by the observation that although the neo-minichromosome carries only one functional centromere, both ends of the minichromosome are heterochromatic, and mouse satellite DNA sequences were found in these heterochromatic regions by *in situ* hybridization.

These two cell lines EC3/7C5 and EC3/7C6, thus, carry a selectable mammalian minichromosome [MMCneo] with a centromere linked to a dominant marker gene Hadlaczký *et al.* (1991) Proc. Natl. Acad. Sci. U.S.A. 88:8106-8110]. MMCneo is intended to be used as a vector for minichromosome-mediated gene transfer and has been used as model of a minichromosome-based vector system.

Long range mapping studies of the MMCneo indicated that human DNA and the *neo* gene constructs integrated into the mouse chromosome separately, followed by the amplification of the chromosome region that contains the exogenous DNA. The MMCneo contains about 30-50 copies of the λ CM8 and λ gtWESneo DNA in the form of approximately 160 kb repeated blocks, which together cover at least a 3.5 Mb region. In addition to these, there are mouse telomeric sequences [Praznovszky *et al.* (1991) Proc. Natl. Acad. Sci. U.S.A. 88:11042-11046] and any DNA of mouse origin necessary for the correct higher-ordered structural organization of chromatids.

Using a chromosome painting probe mCPE1.51 [mouse long interspersed repeated DNA], which recognizes exclusively euchromatic mouse DNA, detectable amounts of interspersed repeat sequences were found on the MMCneo by *in situ* hybridization. The neo-centromere is associated with a small but detectable amount of satellite DNA. The chromosome breakage that separates the neo-centromere from the mouse chromosome occurs in the "foreign" DNA region. This is demonstrated by the presence of lambda and human DNA at the broken end of the formerly dicentric chromosome. At both ends of the MMCneo, however, there traces

of mouse major satellite DNA occur as evidence by *in situ* hybridization. This observation suggests that the doubling in size of the chromosome fragment carrying the neo-centromere during the stabilization of the MMCneo is a result of an inverted duplication. Although mouse telomere sequences, which coamplified with the exogenous DNA sequences during the neo-centromere formation, may provide sufficient telomeres for the MMCneo, the duplication could have supplied the functional telomeres for the minichromosome.

C. Isolation and partial purification of minichromosomes

Mitotic chromosomes of EC3/7C5 cells were isolated as described by Hadlaczký *et al.* [(1981) *Chromosoma* 81:537-555], using a glycine-hexylene glycol buffer system [Hadlaczký *et al.* (1982) *Chromosoma* 86:643-659]. Chromosome suspensions were centrifuged at 1,200 x g for 30 minutes. The supernatant containing minichromosomes was centrifuged at 5,000 x g for 30 minutes and the pellet was resuspended in the appropriate buffer. Partially purified minichromosomes were stored in 50% glycerol at -20° C.

D. Stability of the MMCneo maintenance and *neo* expression

EC3/7C5 cells grown in non-selective medium for 284 days and then transferred to selective medium containing 400 µg/ml G418 showed a 96% plating efficiency (colony formation) compared to control cells cultured permanently in the presence of G418. Cytogenetic analysis indicated that the MMCneo is stably maintained at one copy per cell both under selective and non-selective culture conditions. Only two metaphases with two MMCneo were found in 2,270 metaphases analyzed.

Southern hybridization analysis showed no detectable changes in DNA restriction patterns, and similar hybridization intensities were observed with a *neo* probe when DNA from cells grown under selective or non-selective culture conditions were compared.

Northern analysis of RNA transcripts from the *neo* gene isolated from cells grown under selective and non-selective conditions showed only minor and not significant differences. Expression of the *neo* gene persisted in EC3/7C5 cells maintained in F-12 medium free of G418 for 290 days under non-selective culture conditions. The long term expression of the *neo* gene(s) from the minichromosome may be influenced by the nuclear location of the MMCneo. *In situ* hybridization experiments revealed a preferential peripheral location of the MMCneo in the interphase nucleus. In more than 60% of the 2,500 nuclei analysis, the minichromosome was observed at the perimeter of the nucleus near the nuclear envelope.

EXAMPLE 3

Minichromosome transfer and production of the Λ -neo-chromosome

A. Minichromosome transfer

The neo-minichromosome [referred MMCneo, FIG. 2C] has been used for gene transfer by fusion of minichromosome containing cells [EC3/7C5 or EC3/7C6] with different mammalian cells, including hamster and human. Thirty-seven stable hybrid cell lines have been produced. All established hybrid cell lines proved to be true hybrids as evidenced by *in situ* hybridization using biotinylated human, and hamster genomic, or pMCPE1.51 mouse long interspersed repeated DNA probes for "chromosome painting".

The MMCneo has also been successfully transferred into mouse A9, L929 and pluripotent F9 teratocarcinoma cells by fusion of microcells derived from EC3/7C5 cells. Transfer was confirmed by PCR, Southern blotting and *in situ* hybridization with minichromosome-specific probes. The cytogenetic analysis confirmed that, as expected for microcell fusion, a few cells [1-5%] received [or retained] the MMCneo.

These demonstrate that the MMCneo is tolerated by a wide range of cells. The prokaryotic genes and the extra dosage for the human and lambda sequences carried on the minichromosome seem to be not disadvantageous for tissue culture cells.

The MMCneo is the smallest chromosome of the EC3/7C5 genome and is estimated to be approximately 20-30 Mb, which is significantly smaller than the majority of the host cell (mouse) chromosomes. By virtue of the smaller size, minichromosomes can be partially purified from a suspension of isolated chromosomes by a simple differential centrifugation. In this way, minichromosome suspensions of 15-20% purity have been prepared. These enriched minichromosome preparations can be used to introduce, such as by microinjection or lipofection the minichromosome into selected target cells. Target cells include therapeutic cells that can be used in methods of gene therapy, and also embryonic cells for the preparation of transgenic animals.

The MMCneo is capable of autonomous replication, is stably maintained in cells, and permits persistent expression of the *neo* gene(s), even after long-term culturing under non-selective conditions. It is a non-integrative vector that appears to occupy a territory near the nuclear envelope. Its peripheral localization in the nucleus may have an important role in maintaining the functional integrity and stability of the MMCneo. Functional compartmentalization of the host nucleus may have an effect on the function of foreign sequences. In addition, contains megabases of lambda DNA sequences that should serve as a target site for homologous recombination and thus integration of desired gene(s) into the MMCneo. It can be transferred by cell and microcell fusion, microinjection, electroporation, or chromosome uptake. The neo-centromere of the MMCneo is capable of maintaining and supporting the normal segregation of a larger 150-200 Mb λ neo chromosome. This result (B) demonstrates that the MMCneo chromosome should be useful for carrying large fragments of heterologous DNA.

B. Production of the λ -neo-chromosome

In one hybrid cell line KE1-2/4 made by fusion of EC3/7 and Chinese hamster ovary cells [FIG 2D], the separation of the neo-centromere from the dicentric chromosome was associated with a further amplification process.

- 5 This amplification resulted in the formation of a stable, chromosome of average size [i.e., the λ neo-chromosome; see, Praznovszky et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:11042-11046]. The λ neo-chromosome carries a terminally located functional centromere, and is composed of seven large amplicons containing multiple copies of λ , human, bacterial, and mouse
- 10 DNA sequences [see FIG 2]. The amplicons are separated by mouse major satellite DNA [Praznovszky et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:11042-11046] which forms narrow bands of constitutive heterochromatin between the amplicons.

EXAMPLE 4

- 15 Formation of the "sausage chromosome" [SC]

The findings set forth in the above EXAMPLES demonstrate that the centromeric region of the mouse chromosome 7 has the capacity for large-scale amplification [other results indicate that this capacity is not unique to chromosome 7]. This conclusion is further supported by results from cotransfection experiments, in which a second dominant selectable marker gene and a non-selected marker gene were introduced into EC3/7 cells carrying the formerly dicentric chromosome 7 and the neo-minichromosome.

The EC3/7C5 cell line was transformed with λ phage DNA, a hygromycin construct [pH132], and a β -galactosidase construct [pCH110].

- 25 Stable transformants were selected in the presence of high concentrations [400 $\mu\text{g/ml}$] Hygromycin B, and analysed by Southern hybridization. Established transformant cell lines showing multiple copies of integrated exogenous DNA were studied by *in situ* hybridization to localize the integration site(s), and by LacZ staining to detect β -galactosidase expression.

A. Materials and methods

1. Construction of pH132

The pH132 plasmid carries the hygromycin B resistance gene and the anti-HIV-1 *gag* ribozyme [see, SEQ ID NO. 6 for DNA sequence that corresponds to the sequence of the ribozyme] under control of the β -actin promoter. This plasmid was constructed from pHyg plasmid [Sugden *et al.* (1985) *Mol. Cell. Biol.* 5:410-413; a gift from Dr. A. D. Riggs, Beckman Research Institute, Duarte; see, also, *e.g.*, U.S. Patent No. 4,997,764], and from pPC-RAG12 plasmid [see, Chang *et al.* (1990) *Clin Biotech* 2:23-31; provided by Dr. J. J. Rossi, Beckman Research Institute, Duarte; see, also U.S. Patent Nos. 5,272,262, 5,149,796 and 5,144,019, which describes the anti-HIV *gag* ribozyme and construction of a mammalian expression vector containing the ribozyme insert linked to the β -actin promoter and SV40 late gene transcriptional termination and polyA signals]. The ribozyme insert flanked by BamHI linkers was inserted into BamHI-digested pH β -Apr-1gpt [see, Gunning *et al.* (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84:4831-4835, see, also U.S. Patent No. 5,144,019]. An EcoRI/XhoI fragment of this vector was inserted into EcoRI/XhoI-digested pHyg.

Plasmid pH132 was constructed as follows. First, pPC-RAG12 [described by Chang *et al.* (1990) *Clin. Biotech.* 2:23-31] was digested with BamHI to excise a fragment containing an anti-HIV ribozyme gene [referred to as ribozyme D by Chang *et al.* [(1990) *Clin. Biotech.* 2:23-31]; see also U.S. Patent No. 5,144,019 to Rossi *et al.*, particularly Figure 4 of the patent] flanked by the human β -actin promoter at the 5' end of the gene and the SV40 late transcriptional termination and polyadenylation signals at the 3' end of the gene. As described by Chang *et al.* [(1990) *Clin. Biotech.* 2:23-31], ribozyme D is targeted for cleavage of the translational initiation region of the HIV *gag* gene. This fragment of pPC-RAG12 was subcloned into pBluescript-KS(+) [Stratagene, La Jolla, CA] to produce plasmid 132. Plasmid 132 was then digested with XhoI and EcoRI to yield a fragment

containing the ribozyme D gene flanked by the β -actin promoter at the 5' end and the SV40 termination and polyadenylation signals at the 3' end of the gene. This fragment was ligated to the largest fragment generated by digestion of pHyg [Sugden *et al.* (1985) *Mol. Cell. Biol.* 5:410-413] with *Eco*RI and *S*all to yield pH132. Thus, pH132 is an ~9.3 kb plasmid containing the following elements: the β -actin promoter linked to an anti-HIV ribozyme gene followed by the SV40 termination and polyadenylation signals, the thymidine kinase gene promoter linked to the hygromycin resistance gene followed by the thymidine kinase gene polyadenylation signal, and the *E. coli* ColE1 origin of replication and the ampicillin-resistance gene.

The plasmid pHyg [see, *e.g.*, U.S. Patent Nos. 4,997,764, 4,686,186 and 5,162,215], which confers resistance to hygromycin B using transcriptional controls from the HSV-1 tk gene, was originally constructed from pKan2 [Yates *et al.* (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81:3806-3810] and pLG89 [see, Gritz *et al.* (1983) *Gene* 25:179-188]. Briefly pKan2 was digested with *S*maI and *B*glII to remove the sequences derived from transposon Tn5. The hygromycin-resistance *hph* gene was inserted into the digested pKan2 using blunt-end ligation at the *S*naI site and "sticky-end" ligation [using 1 Weiss unit of T4 DNA ligase (BRL) in 20 microliter volume] at the *B*glII site. The *S*maI and *B*glII sites of pKan2 were lost during ligation.

The resulting plasmid pH132, produced from introduction of the anti-HIV ribozyme construct with promoter and polyA site into pHyg, includes the anti-HIV ribozyme under control of the β -actin promoter as well as the Hyg gene under control of the TK promoter.

2. Chromosome banding

Trypsin G-banding of chromosomes was performed as described in

EXAMPLE 1.

3. Cell cultures

TF1004G19 and TF1004G19C5 mouse cells and the 19xHa4 hybrid, described below, and its sublines were cultured in F-12 medium containing 400 μ g/ml Hygromycin B [Calbiochem].

B. Cotransfection of EC3/7C5 to produce TF1004G-19

- 5 Cotransfection of EC3/7C5 cells with plasmids [pH132, pCH110 available from Pharmacia, see, also Hall *et al.* (1983) *J. Mol. Appl. Gen.* 2:101-109] and with λ DNA [λ cl 875 Sam 7(New England Biolabs)] was made using the calcium phosphate DNA precipitation method [see, *e.g.*, Chen *et al.* (1987) *Mol. Cell. Biol.* 7:2745-2752], using 2-5 μ g plasmid DNA
- 10 and 20 μ g λ phage DNA per 5×10^6 recipient cells.

C. Cell lines containing the sausage chromosome

- One transformant TF1004G-19 was identified. It has a high copy number of integrated pH132 and pCH110 sequences, and a high level of β -galactosidase expression. G-banding and *in situ* hybridization with a human
- 15 probe [CM8; see, *e.g.*, U.S. application Serial No. 08/375,271] revealed unexpectedly that integration had occurred in the formerly dicentric chromosome 7 of the EC3/7C5 cell line. Furthermore, this chromosome carried a newly formed heterochromatic chromosome arm. The size of this heterochromatic arm varied between ~ 150 and ~ 800 Mb in individual
- 20 metaphases.

- By single cell cloning from the TF1004G-19 cell line, a subclone TF1004G-19C5 [FIG 2D], which carries a stable chromosome 7 with a ~ 100 -150 Mb heterochromatic arm [the sausage chromosome] was obtained. This cell line has been deposited in the ECACC under Accession
- 25 No. 96040926. This chromosome arm is composed of four to five satellite segments rich in satellite DNA, and evenly spaced integrated heterologous "foreign" DNA sequences. At the end of the compact heterochromatic arm of the sausage chromosome, a less condensed euchromatic terminal segment is regularly observed. This subclone was used for further
- 30 analyses.

D. Demonstration that the sausage chromosome is derived from the formerly dicentric chromosome

In situ hybridization with lambda phage and pH132 DNA on the TF1004G-19C5 cell line showed positive hybridization only on the minichromosome and on the heterochromatic arm of the "sausage" chromosome [Fig. 2D]. It appears that the "sausage" chromosome [herein also referred to as the SC] developed from the formerly dicentric chromosome (FD) of the EC3/7C5 cell line.

To establish ^{this, the} ~~the, the~~ integration sites of pCH110 and pH132 plasmids was determined. This was accomplished by *in situ* hybridization on these cells with biotin-labeled subfragments of the hygromycin resistance gene and the β -galactosidase gene. Both experiments resulted in narrow hybridizing bands on the heterochromatic arm of sausage chromosome. The same hybridization pattern was detected on the sausage chromosome using a mixture of biotin-labeled λ probe and pH132 plasmid, proving the cointegration of lambda phages, pH132 and pCH110 plasmids.

To examine this further, the cells were cultured in the presence of the DNA-binding dye Hoechst 33258. Culturing of mouse cells in the presence of this dye results in under-condensation of the pericentric heterochromatin of metaphase chromosomes, thereby permitting better observation of the hybridization pattern. Using this technique the heterochromatic arm of sausage chromosome of 19C5 cell showed regular under-condensation revealing the details of the structure of "sausage" chromosome by *in situ* hybridization. Results of *in situ* hybridization on Hoechst-treated TF1004G/19C5 cells with biotin-labeled subfragments of hygromycin resistance and β -galactosidase genes shows that these genes are localized only in the heterochromatic arm of the sausage chromosome. In addition an equal banding hybridization pattern was observed. This pattern of repeating units [amplicons] clearly indicates that the sausage chromosome was formed by an amplification process and that the lambda phage, pH132 and pCH110 plasmid DNA sequences border the amplicons.

In another series of experiments, using fluorescence *in situ* hybridization [FISH] was carried out with mouse major satellite DNA, the main component of the mouse pericentric heterochromatin, the results confirmed that the amplicons of the sausage chromosome are primarily composed of satellite DNA.

E. The sausage chromosome has one centromere

To determine whether mouse centromeric sequences had participated in the amplification process forming the "sausage" chromosome and whether or not the amplicons carry inactive centromeres, *in situ* hybridization was carried out with mouse minor satellite DNA. Mouse minor satellite DNA is localized specifically near the centromeres of all mouse chromosome. Positive hybridization was detected in all mouse centromeres including the sausage chromosome, which, however, only showed a positive signal at the beginning of the heterochromatic arm.

Indirect immunofluorescence with human anti-centromere antibody [LU 851] that only which can recognizes the functional centromeres [see, e.g., Hadlaczký *et al.* (1989) *Chromosoma* 97:282-288] proved that sausage chromosome has only one active centromere. The centromere comes from the formerly dicentric part of the chromosome and co-localizes with the *in situ* hybridization signal of the mouse minor DNA probe.

F. The selected and non-selected heterologous DNA in the heterochromatin of the sausage chromosome is expressed

1. High levels of the heterologous genes are expressed

The TF1004G/19C5 cell line thus carries multiple copies of hygromycin resistance and β -galactosidase genes localized only in the heterochromatic arm of the sausage chromosome. The 19C5 cells can grow very well in the presence of 200 μ g/ml or even 400 μ g/ml hygromycin B. (The level of expression was determined by Northern hybridization with subfragment of hygromycin resistance gene and single copy gene).

The expression of the non-selected β -galactosidase gene in the TF1004G/19C5 transformant, was detected with LacZ staining of the cells.

By this method one hundred percent of the cells stained dark blue, showing that there is a high level of β -galactosidase expression in all of 19C5 cells.

2. The heterologous genes that are expressed are in the heterochromatin

5 To demonstrate that the genes localized in the constitutive heterochromatin of the sausage chromosome provide the hygromycin resistance and the LacZ staining capability of 19C5 transformant [i.e. β -gal express], PEG induced cell fusion between a TF1004G/19C5 mouse cell and Chinese hamster ovary cell was performed. The hybrids were selected and
10 maintained in HAT medium containing G418 [400 μ g/ml] and hygromycin [200 μ g/ml]. Two hybrid clones designated 19C5xHa3 and 19C5xHa4, which has been deposited in the ECACC under Accession No. 96040927, were selected. Both carry the sausage chromosome and the minichromosome.

15 Twenty-seven single cell derived colonies of 19C5xHa4 hybrid were maintained and analyzed as individual subclones. *In situ* hybridization with hamster and mouse chromosome painting probes and hamster chromosome 2 specific probes verified that the 19C5xHa4 clone contains the complete Chinese hamster genome and a partial mouse genome. All 19C5xHa4
20 subclones retained the hamster genome, but different subclones showed different number of mouse chromosomes indicating the preferential elimination of mouse chromosomes.

To promote further elimination of mouse chromosomes hybrid cells were repeatedly treated with BrdU. The BrdU treatments, which destabilize
25 the genome, result in significant loss of mouse chromosomes. The BrdU treated 19C5xHa4 hybrid cells were divided to three groups. One group of the hybrid cells (GH)^{was} maintained in the presence of hygromycin (200 μ g/ml) and G418 (400 μ g/ml), and the other two groups of the cells were
a cultured under G418 (G) or hygromycin (H) selection conditions to promote
30 the elimination of sausage or minichromosome.

One month later, single cell derived subclones were established from these three subcultures of the TF1004G/19C5xHa4 hybrid line. The subclones were monitored by *in situ* hybridization with biotin-labeled lambda phage and hamster chromosome painting probes. Four individual clones [G2B5, G3C5, G4D6, G2B4] selected in the presence of G418 that have lost the sausage chromosome but retained minichromosome were found. Under hygromycin selection only one subclone [H1D3] lost the minichromosome. In this clone the sausage chromosome was present.

Since hygromycin resistance and β -galactosidase genes were thought to be expressed from the sausage chromosome, the expression of these genes were analyzed in the four subclones that had lost the sausage chromosome. In the presence of 200 μ g/ml hygromycin, one hundred percent of the cells of four individual subclones died. In order to detect the β -galactosidase expression hybrid subclones were analyzed by LacZ staining.

One hundred percent of the cells of the four subclones that lost the sausage chromosome also lost the LacZ staining capability. All of the other hybrid subclones that not lost the sausage chromosome under the non-selective culture conditions showed positive LacZ staining.

These findings demonstrate that the expression of hygromycin resistance and β -galactosidase genes is linked to the presence of the sausage chromosome and results of *in situ* hybridizations show that the heterologous DNA is expressed from the constitutive heterochromatin of the sausage chromosome.

By *in situ* hybridization in three other hybrid subclones [G2C6, G2D1, and G4D5] the presence of the sausage chromosome was not detected. By the LacZ staining method some stained cells were detected in these hybrid lines and when these subclones were transferred to hygromycin selection some colonies survived. Cytological analysis and *in situ* hybridization of these hygromycin resistance colonies revealed the presence of the sausage chromosome, suggesting that only the cells of G2C6, G2D1 and G4D5

hybrids that had not lost the sausage chromosome were able to preserve the hygromycin resistance and β -galactosidase expression. These results confirmed that the expression of these genes is linked to the presence of the sausage chromosome. The level of β -galactosidase expression was

5 determined by the immunoblot technique using monoclonal antibody.

Hygromycin resistance and β -galactosidase expression of the cells which contained the sausage chromosome were provided by the genes localized in the mouse pericentric heterochromatin. This was demonstrated by performing Southern DNA hybridizations on the hybrid cells that lack the

10 sausage chromosome with PCR amplified subfragments of hygromycin resistance and β -galactosidase genes. None of the subclones showed hybridization with these probes, however, all of the analyzed clones contained the minichromosome as well. Other hybrid clones that contain the sausage chromosome showed intense hybridization with these DNA probes.

15 These results lead to the conclusion that hygromycin resistance and β -galactosidase expression of the cells that contain the sausage chromosome were provided by the genes localized in the mouse pericentric heterochromatin.

EXAMPLE 5

20 The gigachromosome

As described in Example 4, the sausage chromosome was transferred into Chinese hamster cells by cell fusion. Using Hygromycin B/HAT selection, two hybrid clones 19C5xHa3 and 19C5xHa4 were produced that carry sausage chromosome. *In situ* hybridization, using hamster and mouse

25 chromosome-painting probes and a hamster chromosome 2 specific probe, verified that clone 19C5xHa4 contains a complete Chinese hamster genome as well as partial mouse genomes. Twenty-seven separate colonies of 19C5xHa4 cells were maintained and analyzed as individual subclones. Twenty-six out of 27 subclones contained a morphologically unchanged

30 sausage chromosome.

In one clone 19C5xHa47 [see Fig. 2E], the heterochromatic arm of the sausage chromosome became unstable and showed continuous intrachromosomal growth. In extreme cases, the amplified chromosome arm exceeded 1000 Mb in size (gigachromosome).

EXAMPLE 6

5

The stable megachromosome

A. Formation of the megachromosome

- All 19C5xHa4 subclones retained a complete hamster genome, but different subclones showed different numbers of mouse chromosomes, indicating the preferential elimination of mouse chromosomes. As described in Example 4, to promote further elimination of mouse chromosomes, hybrid cells were repeatedly treated with 10^{-4} M BrdU for 16 hours and single cell subclones were established. The BrdU treatments appeared to destabilize the genome, resulting in a change in the sausage chromosome as well. A gradual increase in a cell population in which a further amplification had occurred was observed. In addition to the ~100-150 Mb heterochromatic arm of the sausage chromosome, an extra centromere and a ~150-250 Mb heterochromatic chromosome arm were formed, which differed from those of mouse chromosome 7. By the acquisition of another euchromatic terminal segment, a new submetacentric chromosome (megachromosome) was formed. Seventy-nine individual subclones were established from these BrdU-treated cultures, by single-cell cloning: 42 subclones carried the intact megachromosome, 5 subclones carried the sausage chromosome, and in 32 subclones fragments or translocated segments of the megachromosome were observed. Twenty-six subclones were cultured under non-selective conditions over a 2 month period. In 19 out of 26 subclones, the megachromosome was retained. Those subclones which lost the megachromosomes all became sensitive to Hygromycin B and had no β -galactosidase expression, indicating that both markers were linked to the megachromosome.

Two sublines (G3D5 and H1D3), which were chosen for further experiments, showed no changes in the morphology of the megachromosome during more than 100 generations under selective conditions.

5 B. Structure of the megachromosome

The following results demonstrate that, apart from the euchromatic terminal segments, the whole megachromosome is constitutive heterochromatin, containing a tandem array of at least 40 [~ 7.5 Mb] blocks of mouse major satellite DNA [see Figures 2 and 3]. Four satellite DNA blocks are organized into a giant palindrome [amplicon] carrying integrated exogenous DNA sequences at each end. The long and short arms of the submetacentric megachromosome contains 6 and 4 amplicons, respectively.

15 1. The megachromosome is composed primarily of heterochromatin

Except for the terminal regions, the megachromosome is composed primarily of heterochromatin. This was demonstrated by G-banding of the megachromosome,, which resulted in positive staining characteristic of constitutive heterochromatin. Apart from the terminal regions, the whole megachromosome appears to be heterochromatic. Mouse major satellite DNA is the main component of the pericentric, constitutive heterochromatin of mouse chromosomes and represents $\sim 10\%$ of the total DNA [Waring *et al.* (1966) *Science* 154:791-794]. Using a mouse major satellite DNA probe for *in situ* hybridization, strong hybridization was observed throughout the megachromosome, except for its terminal regions. The hybridization showed a segmented pattern: four large blocks appeared on the short arm and usually 4-7 blocks were seen on the long arm. By comparing these segments with the pericentric regions of normal mouse chromosomes that carry approximately ~ 15 Mb of major satellite, the size of the blocks of major satellite on the megachromosome was estimated to be ~ 30 Mb.

Using a mouse probe specific to euchromatin [pMCPE1.51; a mouse long interspersed repeated DNA probe], positive hybridization was detected only on the terminal segments of the megachromosome of the H1D3 hybrid subline. In the G3D5 hybrids, hybridization with a hamster-specific probe revealed that several megachromosomes contained terminal segments of hamster origin on the long arm. This observation indicated that the acquisition of the terminal segments on these chromosomes happened in the hybrid cells, and that the long arm of the megachromosome was the recently formed one.

When a mouse minor satellite probe was used, specific to the centromeres of mouse chromosomes [Wong et al. (1988) Nucl. Acids Res. 16:11645-11661], a strong hybridization signal was detected only at the primary constriction of the megachromosome, which colocalized with the positive immunofluorescence signal produced with human anti-centromere serum [LU 851].

In situ hybridization experiments with pH132, pCH110, and λ DNA probes revealed that all heterologous DNA was located in the gaps between the mouse major satellite segments. Each segment of mouse major satellite was bordered by a narrow band of integrated heterologous DNA, except at the second segment of the long arm where a double band existed, indicating that the major satellite segment was missing or considerably reduced in size here. This chromosome region served as a useful cytological marker in identifying the long arm of the megachromosome. At a frequency of 10^{-4} , "restoration" of these missing satellite DNA blocks was observed in one chromatid, when the formation of a whole segment on one chromatid occurred.

After Hoechst 33258 treatment ($50 \mu\text{g/ml}$ for 16 hours), the megachromosome showed undercondensation throughout its length except for the terminal segments. This made it possible to study the architecture of the megachromosome at higher resolution. *In situ* hybridization with the mouse major satellite probe on undercondensed megachromosomes

demonstrated that the ~30 Mb major satellite segments were composed of four blocks of ~7.5 Mb separated from each other by a narrow band of non-hybridizing sequences [Figure 3]. Similar segmentation can be observed in the large block of pericentric heterochromatin in metacentric mouse chromosomes from the LMTK⁻ and A9 cell lines.

2. The megachromosome is composed of segments containing two tandem ~7.5 Mb blocks are followed by two inverted blocks

Because of the asymmetry in thymidine content between the two strands of the DNA of the mouse major satellite, when mouse cells are grown in the presence of BrdU for a single S phase, the constitutive heterochromatin shows lateral asymmetry after FPG staining. Also, in the 19C5xHa4 hybrids, the thymidine-kinase the [*Tk*] deficiency of the mouse fibroblast cells was complemented by the hamster *Tk* gene, permitting BrdU incorporation experiments.

A striking structural regularity in the megachromosome was detected using the FPG technique. In both chromatids, alternating dark and light staining that produced a checkered appearance of the megachromosome was observed. A similar picture was obtained by labelling with fluorescein-conjugated anti-BrdU antibody. Comparing these pictures to the segmented appearance of the megachromosome, showed that one dark and one light FPG band corresponded to one ~30 Mb segment of the megachromosome.

These results suggest that the two halves of the ~30 Mb segment have an inverted orientation. This was verified by combining *in situ* hybridization and immunolabelling of the incorporated BrdU with fluorescein-conjugated anti-BrdU antibody on the same chromosome. Since the ~30 Mb segments of the megachromosome are composed of four blocks of mouse major satellite, it can be concluded that two tandem ~7.5 Mb blocks are followed by two inverted blocks within one segment.

Large-scale mapping of megachromosome DNA by pulsed-field electrophoresis and Southern hybridization with "foreign" DNA probes revealed a simple pattern of restriction fragments. Using endonucleases with none, or only a single cleavage site in the integrated foreign DNA sequences, followed by hybridization with a *hyg* probe, 1-4 predominant fragments were detected. Since the megachromosome contains 10-12 amplicons with an estimated 3-8 copies of *hyg* sequences per amplicon (30-90 copies per megachromosome), the small number of hybridizing fragments indicates the homogeneity of DNA in the amplified segments.

10 3. Scanning electron microscopy of the megachromosome confirmed the above findings

The homogeneous architecture of the heterochromatic arms of the megachromosome was confirmed by high resolution scanning electron microscopy. Extended arms of megachromosomes, and the pericentric heterochromatic region of mouse chromosomes, treated with Hoechst 33258, showed similar structure. The constitutive heterochromatic regions appeared more compact than the euchromatic segments. Apart from the terminal regions, both arms of the megachromosome were completely extended, and showed faint grooves, which should correspond to the border of the satellite DNA blocks in the non-amplified chromosomes and in the megachromosome. Without Hoechst treatment, the grooves seemed to correspond to the amplicon borders on the megachromosome arms. In addition, centromeres showed a more compact, finely fibrous appearance than the surrounding heterochromatin.

25 C. Formation of the megachromosome

Figure 2 schematically sets forth the events leading to the formation of the stable megachromosome: (A) A single E-type amplification in the centromeric region of chromosome 7 generates the neo-centromere linked to the integrated foreign DNA, and forms a dicentric chromosome. Multiple E-type amplification forms the λ neo-chromosome, which was derived from chromosome 7 and stabilized in a mouse-hamster hybrid cell line; (B) Specific

breakage between the centromeres of a dicentric chromosome 7 generates a chromosome fragment with the neo-centromere, and a chromosome 7 with traces of foreign DNA at the end; (C) Inverted duplication of the fragment bearing the neo-centromere results in the formation of a stable neo-minichromosome; (D) Integration of exogenous DNA into the foreign DNA region of the formerly dicentric chromosome 7 initiates H-type amplification, and the formation of a heterochromatic arm. By capturing a euchromatic terminal segment, this new chromosome arm is stabilized in the form of the "sausage" chromosome; (E) BrdU treatment and/or drug selection appears to induce further H-type amplification, which results in the formation of an unstable gigachromosome: (F) Repeated BrdU treatments and/or drug selection induce further H-type amplification including a centromere duplication, which leads to the formation of another heterochromatic chromosome arm. It is split off from the chromosome 7 by chromosome breakage and acquires a terminal segment to form the stable megachromosome.

EXAMPLE 7

Summary of some of the cell lines with SATACS and minichromosomes that have been constructed

have been constructed

20 LMTK⁻-derived cell line, which is a mouse fibroblast cell line, was transfected with λ CM8 and λ gtWESneo DNA [see, EXAMPLE 2] to produce transformed cell lines. Among these cell lines was EC3/7, deposited at the European Collection of Animal cell Culture (ECACC) under Accession No. 90051001 [see, U.S. Patent No. 5,288,625; see, also Hadlaczky et al.

25 (1991) Proc. Natl. Acad. Sci. U.S.A. 88:8106-8110 and U.S. application Serial No. 08/375,271]. This cell line contains the dicentric chromosome with the neo centromere. Recloning and selection produced cell lines such as EC3/7C5, which are cell lines with the stable neo-minichromosome [see, Fig. 2C].

Fig. 2C].

30 Fusion of EC3/7 with CHO-K20 cells and selection with G418/HAT produced hybrid cell lines, among these was KE1 2/4, which has been

deposited with the ECACC under Accession No. 96040924. KE1 2/4 is a stable cell line that contains the λ neo-chromosome [see, Fig. 2D; see, also U.S. Patent No. 5,288,625], produced by E type amplifications. KE1 2/4 has been transfected with vectors containing lambda DNA, selectable markers, such as puromycin resistance, and genes of interest, such as p53, anti-HIV ribozymes. These vectors target the gene of interest into the λ neo-chromosome by virtue of homologous recombination with the heterologous DNA in the chromosome.

The EC3/7C5 cell line has been co-transfected with pH132, pCH110 and λ DNA [see, EXAMPLE 2] as well as other constructs. Various clones and subclones have been selected. For example transformation with a construct that includes p53 encoding DNA, produced cells designated C5pMCT53.

As discussed above, cotransfection of EC3/7C5 cells with plasmids [pH132, pCH110 available from Pharmacia, see, also Hall *et al.* (1983) *J. Mol. Appl. Gen.* 2:101-109] and with λ DNA [λ cl 875 Sam 7(New England Biolabs)] produced transformed cells. Among these is TF1004G24, which contains the DNA encoding the anti-HIV ribozyme in the neo-minichromosome. Recloning of TF1004G24 produced numerous cell lines. Among these the NHHL24 cell line. This cell line also has the anti-HIV ribozyme in the neo-minichromosome and expresses high levels of β -gal. It has been fused with CHO-K20 cells to produce various hybrids.

Recloning and selection of the TF1004G transformants produced the cell line TF1004G-19, discussed above in EXAMPLE 4, which contains the unstable sausage chromosome. Single cell cloning produced the TF1004G-19C5 [see Figure 4] cell line, which has a stable sausage chromosome. TF1004G-19C5 has been fused with CHO cells and the hybrids grown under selective conditions to produce the 19C5xHa4 cell line [see, EXAMPLE 4] and others. BrdU treatment of 19C5xHa4 cells and growth under selective conditions [neomycin (G) and/or hygromycin (H)] ^{has} produced

hybrid cell lines with the gigachromosome [see Figure 2E] and the G3D5 and G4D6 cell lines and others. G3D5 has the neo-minichromosome and the megachromosome, G4D6 has only the neo-minichromosome.

- 5 Recloning of G3D5 in GH medium produced numerous clones. Among these is H1D3 [see Figure 4], which has the stable megachromosome. Repeated BrDU treatment and recloning has produced the HB31 cell line, which has been used for transformations with the pTEMPUD, pTEMPU and pTEMPU3 vectors [see, Example 12, below].

- 10 H1D3 has been fused with a CD4⁺ Hela cell line that carries DNA encoding CD4 and neomycin resistance on a plasmid [see, e.g., U.S. Patent Nos. 5,413,914, 5,409,810, 5,266,600, 5,223,263, 5,215,914 and 5,144,019, which describe these Hela cells]. Selection with GH has produced hybrids, including H1xHE41 [see Figure 4], which carries the megachromosome and also a single human chromosome that includes the
- 15 CD4neo construct. Repeated BrdU treatment and single cell cloning has produced cell lines with the megachromosome [cell line 1B3, see Figure 4], cell lines, such as 1B4 and others that have a dwarf megachromosome [~150-200 Mb] and cell lines, such as 1C3, which has micro-megachromosome [~60-90 Mb]. About 25% of the 1B3 cells have a
- 20 truncated megachromosome [~90-120 Mb].

EXAMPLE 8

Replication of the megachromosome

- This homogeneous architecture of the megachromosomes provides a unique opportunity to perform a detailed analysis of the replication of the
- 25 constitutive heterochromatin.

A. Materials and methods

1. Culture of cell lines

- H1D3 mouse-hamster hybrid cells carrying the megachromosome [see, EXAMPLE 4] were cultured in F-12 medium containing 10% fetal calf serum
- 30 [FCS] and 400 μ g/ml Hygromycin B [Calbiochem]. G3D5 hybrid cells [see,

Example 4] were maintained in F-12 medium containing 10% FCS, 400 $\mu\text{g/ml}$ Hygromycin B (Calbiochem), and 400 $\mu\text{g/ml}$ G418[SIGMA]. Mouse A9 fibroblast cells were cultured in F-12 medium supplemented with 10% FCS.

2. BrdU labelling

- 5 In typical experiments, 20-24 parallel semi-confluent cell cultures were set up in 10 cm Petri dishes. Bromodeoxyuridine (BrdU) (Fluka) was dissolved in distilled water alkalized with a drop of NaOH, to make a 10^{-2}M stock solution. Aliquots of 10-50 μl of this BrdU stock solution were added to each 10 ml culture, to give a final BrdU concentration of 10-50 μM . The
- 10 cells were cultured in the presence of BrdU for 30 min, and then washed with warm complete medium, and incubated without BrdU until required. At this point, 5 $\mu\text{g/ml}$ colchicine was added to a sample culture every 1 or 2 h. After 1-2 h colchicine treatment, mitotic cells were collected by "shake-off" and regular chromosome preparations were made for immunolabelling.

3. Immunolabelling of chromosomes and *in situ* hybridization

- 15 Immunolabelling with fluorescein-conjugated anti-BrdU monoclonal antibody (Boehringer) was done according to the manufacturer's recommendations, except that for mouse A9 chromosomes, 2 M hydrochloric acid was used at 37°C for 25 min, while for chromosomes
- 20 of hybrid cells, 1 M hydrochloric acid was used at 37°C for 30 min. *In situ* hybridization with biotin-labelled probes, and indirect immunofluorescence and *in situ* hybridization on the same preparation, were performed as described previously [Hadlaczký *et al.* (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:8106-8110, see, also U.S. Patent No. 5,288,625].

4. Microscopy

- 25 All observations and microphotography were made by using a Vanox AHBS (Olympus) microscope. Fujicolor 400 Super G or Fujicolor 1600 Super HG high-speed colour negatives were used for photographs.

B. Results

The replication of the megachromosome was analyzed by BrdU pulse labelling followed by immunolabelling. The basic parameters for DNA labelling *in vivo* were first established. Using a 30 min pulse of 50 μ M BrdU in parallel cultures, samples were taken and fixed at 5 min intervals from the beginning of the pulse, and every 15 min up to 1 h after the removal of BrdU. Incorporated BrdU was detected by immunolabelling with fluorescein-conjugated anti-BrdU monoclonal antibody. At the first time point (5 min) 38% of the nuclei were labelled, and a gradual increase in the number of labelled nuclei was observed during incubation in the presence of BrdU, culminating in 46% in the 30 min sample, at the time of the removal of BrdU. At further time points (60, 75, and 90 min) no significant changes were observed, and the fraction of labelled nuclei remaining constant [44.5-46%].

These results indicate that (i) the incorporation of the BrdU is a rapid process, (ii) the 30 min pulse-time is sufficient for reliable labelling of S-phase nuclei, and (iii) the BrdU can be effectively removed from the cultures by washing.

The length of the cell cycle of the 19C5xHa4 hybrid cells was estimated by measuring the time between the appearance of the earliest BrdU signals on the extreme late replicating chromosome segments and the appearance of the same pattern only on one of the chromatids of the chromosomes after one completed cell cycle. The length of G2 period was determined by the time of the first detectable BrdU signal on prophase chromosomes and by the labelled mitoses method [Qastler *et al.* (1959) *Exp. Cell Res.* 17:420-438]. The length of the S-phase was determined in three ways: (i) on the basis of the length of cell cycle and the fraction of nuclei labelled during the 30-120 min pulse; (ii) by measuring the time between the very end of the replication of the extreme late replicating chromosomes and the detection of the first signal on the chromosomes at the beginning of S phase; (iii) by the labelled

mitoses method. In repeated experiments, the duration of the cell cycle was found to be 22-26 h, the S phase 10-14 h, and the G2 phase 3.5-4.5 h.

- Analyses of the replication of the megachromosome were made in parallel cultures by collecting mitotic cells at two hour intervals following two hours of colchicine treatment. In a repeat experiment, the same analysis was performed using one hour sample intervals and one hour colchicine treatment. Although the two procedures gave comparable results, the two hour sample intervals were viewed as more appropriate since approximately 30% of the cells were found to have a considerably shorter or longer cell cycle than the average. The characteristic replication patterns of the individual chromosomes, especially some of the late replicating hamster chromosomes, served as useful internal markers for the different stages of S-phase. To minimize the error caused by the different lengths of cell cycles in the different experiments, samples were taken and analyzed throughout the whole cell cycle until the appearance of the first signals on one chromatid at the beginning of the second S-phase.

- The sequence of replication in the megachromosome is as follows.
- At the very beginning of the S-phase, the replication of megachromosome starts at the ends of the chromosomes. The first initiation of replication in an interstitial position can usually be detected at the centromeric region. Soon after, but still in the first quarter of the S-phase, when the terminal region of the short arm has almost completed its replication, discrete initiation signals appear along the chromosome arms. In the second quarter of the S-phase, as replication proceeds, the BrdU-labelled zones gradually widen, and the checkered pattern of the megachromosome becomes clear [see, e.g., Fig. 2F]. At the same time, pericentric regions of mouse chromosomes also show intense incorporation of BrdU. The replication of the megachromosome peaks at

the end of the second quarter and in the third quarter of the S-phase. At the end of the third quarter, and at the very beginning of the last quarter of the S-phase, the megachromosome and the pericentric heterochromatin of the mouse chromosomes complete their replication. By the end of S-
5 phase only the very late replicating segments of mouse and hamster chromosomes are still incorporating BrdU.

The replication of the whole genome occurs in distinct phases. The signal of incorporated BrdU increased continuously until the end of the first half of the S-phase, but at the beginning of the third quarter of the S-
10 phase chromosome segments other than the heterochromatic regions hardly incorporated BrdU. In the last quarter of the S-phase, the BrdU signals increased again when the extreme late replicating segments showed very intense incorporation.

Similar analyses of the replication in mouse A9 cells were
15 performed as controls. To increase the resolution of the immunolabelling pattern, pericentric regions of A9 chromosomes were decondensed by treatment with Hoechst 33258. Because of the intense replication of the surrounding euchromatic sequences, precise localization of the initial BrdU signal in the heterochromatin was normally difficult, even on
20 undercondensed mouse chromosomes. On those chromosomes where the initiation signal(s) were localized unambiguously, the replication of the pericentric heterochromatin of A9 chromosomes was similar to that of the megachromosome. Chromosomes of A9 cells also exhibited replication patterns and sequences similar to those of the mouse chromosomes in
25 the hybrid cells. These results indicate that the replicators of the megachromosome and mouse chromosomes retained their original timing and specificity in the hybrid cells.

By comparing the pattern of the initiation sites obtained after BrdU incorporation with the location of the integration sites of the "foreign"
30 DNA in a detailed analysis of the first quarter of the S-phase, an attempt

was made to identify origins of replication (initiation sites) in relation to the amplicon structure of the megachromosome. The double band of integrated DNA on the long arm of the megachromosome served as a cytological marker. The results showed a colocalization of the BrdU and *in situ* hybridization signals found at the cytological level, indicating that the "foreign" DNA sequences are in close proximity to the origins of replication, presumably integrated into the non-satellite sequences between the replicator and the satellite sequences [see, Figure 3]. In the pericentric region of several other chromosomes, dot-like BrdU signals can also be observed that are comparable to the initiation signals on the megachromosome. These signals may represent similar initiation sites in the heterochromatic regions of normal chromosomes.

At a frequency of 10^{-4} , "uncontrolled" amplification of the integrated DNA sequences was observed in the megachromosome.

Consistent with the assumption (above) that "foreign" sequences are in the proximity to the replicators, this spatially restricted amplification is likely to be a consequence of uncontrolled repeated firings of the replication origin(s) without completing the replication of the whole segment.

C. Discussion

It has generally been thought that the constitutive heterochromatin of the pericentric regions of chromosomes is late replicating [see, e.g., Miller (1976) *Chromosoma* 55:165-170]. On the contrary, these experiments evidence that the replication of the heterochromatic blocks starts at a discrete initiation site in the first half of the S-phase and continues through approximately three-quarters of S-phase. This difference can be explained in the following ways: (i) in normal chromosomes, actively replicating euchromatic sequences that surround the satellite DNA obscure the initiation signals, and thus the precise localization of initiation sites is obscured; (ii) replication of the

heterochromatin can only be detected unambiguously in a period during the second half of the S-phase, when the bulk of the heterochromatin replicates and most other chromosomal regions have already completed their replication, or have not yet started it. Thus, low resolution

5 cytological techniques, such as analysis of incorporation of radioactively labelled precursors by autoradiography, only detect prominent replication signals in the heterochromatin in the second half of S-phase, when adjacent euchromatic segments are no longer replicating.

In the megachromosome, the primary initiation sites of replication

10 colocalize with the sites where the "foreign" DNA sequences are integrated at the amplicon borders. Similar initiation signals were observed at the same time in the pericentric heterochromatin of mouse chromosomes that do not have "foreign" DNA, indicating that the replication initiation sites at the borders of amplicons may reside in the

15 non-satellite flanking sequences of the satellite DNA blocks. The presence of a primary initiation site at each satellite DNA doublet implies that this large chromosome segment is a single huge unit of replication [megareplicon] delimited by the primary initiation site and the termination point at each end of the unit. Several lines of evidence indicate that,

20 within this higher-order replication unit, "secondary" origins and replicons contribute to the complete replication of the megareplicon:

1. The total replication time of the heterochromatic regions of the megachromosome was ~9-11 h. At the rate of movement of replication forks, 0.5-5 kb per minute, that is typical of eukaryotic chromosomes
- 25 [Kornberg et al. (1992) *DNA Replication*. 2nd. ed., New York: W.H. Freeman and Co, p. 474], replication of a ~15 Mb replicon would require 50-500 h. Alternatively, if only a single replication origin was used, the average replication speed would have to be 25 kb per minute to complete replication within 10 h. By comparing the intensity of the BrdU signals on
- 30 the euchromatic and the heterochromatic chromosome segments, no

evidence for a 5- to 50-fold difference in their replication speed was found.

2. Using short BrdU pulse labelling, a single origin of replication would produce a replication band that moves along the replicon, reflecting the movement of the replication fork. In contrast, a widening of the replication zone that finally gave rise to the checkered pattern of the megachromosome and within the replication period the most intensive BrdU incorporation occurred in the second half of the S-phase was observed. This suggests that once the megareplicator has been activated, it permits the activation and firing of "secondary" origins, and that the replication of the bulk of the satellite DNA takes place from these "secondary" origins during the second half of the S-phase. This is supported by the observation that in certain stages of the replication of the megachromosome, the whole amplicon can apparently be labelled by a short BrdU pulse.

- Megareplicators and secondary replication origins seem to be under strict temporal and spatial control. The first initiation within the megachromosomes usually occurred at the centromere, and that shortly afterward all the megareplicators become active. The last segment of the megachromosome to complete replication was usually the second segment of the long arm. Results of control experiments with mouse A9 chromosomes indicate that replication of the heterochromatin of mouse chromosomes corresponds to the replication of the megachromosome amplicons. Therefore, the pre-existing temporal control of replication in the heterochromatic blocks is preserved in the megachromosome. Positive [Hassan *et al.* (1994) *J. Cell. Sci.* 107:425-434] and negative [Haase *et al.* (1994) *Mol. Cell. Biol.* 14:2516-2524] correlations between transcriptional activity and initiation of replication have been proposed. In the megachromosome, transcription of the integrated genes seems to have no effect on the original timing of the replication origins. The

of replication.

Considering that pericentric heterochromatin of mouse chromosomes contains thousands of short, simple repeats spanning 7-15 Mb, and the centromere itself may also contain hundreds of kilobases, the existence of a higher-order unit of replication seems probable. The observed uncontrolled intrachromosomal amplification restricted to a replication initiation region of the megachromosome is highly suggestive of a rolling-circle type amplification, and provides additional evidence for the presence of a replication origin in this region.

The finding that a specific replication initiation site occurs at the boundaries of amplicons suggests that replication might play a role in the amplification process. These results suggest that each amplicon of the megachromosome can be regarded as a huge megareplicon defined by a primary initiation site [megareplicator] containing "secondary" origins of replication. Fusion of replication bubbles from different origins of bi-directional replication [DePamphilis (1993) Ann. Rev. Biochem. 62:29-63] within the megareplicon could form a giant replication bubble, which would correspond to the whole megareplicon. In the light of this, the formation of megabase-size amplicons can be accommodated by a replication-directed amplification mechanism. In both H and E-type amplifications, intrachromosomal multiplication of the amplicons was observed [see, above EXAMPLES], which is consistent with the unequal sister chromatid exchange model. Induced or spontaneous unscheduled replication of a megareplicon in the constitutive heterochromatin may also form new amplicon(s) leading to the expansion of the amplification or to the heterochromatic polymorphism of "normal" chromosomes. The "restoration" of the missing segment on the long arm of the

Taken together, without being bound by any theory, a replication-directed mechanism is a plausible explanation for the initiation of large-scale amplifications in the centromeric regions of mouse chromosomes, as well as for the *de novo* chromosome formations. If specific [amplificator] sequences play a role in promoting the amplification process, sequences at the primary replication initiation site [megareplicator] of the megareplicon are possible candidates.

EXAMPLE 9

20 To show that the events described in EXAMPLES 2-7 are not unique to mouse chromosome 7 and to show that the EC7/3 cell line is not required for formation of these chromosomes, the experiments have been repeated using different initial cell lines and DNA fragments. Any cell or cell line should be amenable to use or can readily be determined to be amenable or not.

A. Materials

The LP11 cell line was produced by the "scrape-loading" transfection method [Fechheimer et al. (1987) Proc. Natl. Acad. Sci. U.S.A. 84:8463-8467] using 25 μ g plasmid DNA for 5×10^6 recipient cells. LP11 cells were maintained in F-12 medium containing 3-15 μ g/ml Puromycin [SIGMA].

30 B. **Amplification in EFT1 cells**
 The large-scale amplification described in the above Examples is not restricted to the transformed EC3/7 cell line or to the chromosome 7

- of mouse. In an independent transformation experiment, using a selectable puromycin construct pPuroTel, an LMTK⁻ cell line [LP11] was established, carrying chromosome(s) with amplified chromosome segments of different lengths [~ 150 -600 Mb]. Cytological analysis of
- 5 the LP11 cells indicated that the amplification occurred in the pericentric region of the long arm of a submetacentric chromosome formed by Robertsonian translocation. This chromosome arm was identified by G-banding as chromosome 1. C-banding and *in situ* hybridization with
- 10 mouse major satellite DNA probe showed that an E-type amplification had occurred: the newly formed region was composed of an array of euchromatic chromosome segments containing different amounts of heterochromatin. The size and C-band pattern of the amplified segments were heterogeneous. In several cells, the number of these amplified units
- 15 stable marker chromosomes with 10-15 segments and constant C-band patterns.

EXAMPLE 10

Purification of artificial chromosomes

I. Cell Sorting based on base composition and size

20 A. Cell Lines

- 1B3 mouse-hamster-human hybrid cells [see, Figure 4] carrying the megachromosome or the truncated megachromosome were grown in F-12 medium supplemented with 10% fetal calf serum, 150 μ g/ml hygromycin B and 400 μ g/ml G418. GHB-42 [a cell line recloned from G3D5] mouse-
- 25 hamster hybrid cells carrying the megachromosome and the mini-chromosome were cultured in F-12 medium containing 10% fetal calf serum, 150 μ g/ml hygromycin B and 400 μ g/ml G418. The doubling time of both cell lines was about 24 hours.

B. Chromosome Isolation

To accumulate mitotic cells, 5 μ g/ml colchicine was added for 12 hours to the cultures. Mitotic cells were then harvested by gentle pipetting of the medium on the layer cells. The mitotic index obtained was 60-80%. The mitotic cells by were collected by selective detachment. The cells were sedimented by centrifugation of 200 g for 10 minutes.

Two procedures were used to prepare metaphase chromosomes from these cells, one based on polyamine buffer system. [Cram *et al.* (1990) Methods in Cell Biology 33:377-382] and the other on modified hexylene glycol buffer system [Hadlaczky *et al.* (1982) Chromosoma 86:643-65].

1. Polyamine procedure

In the polyamine procedure, about 10^7 mitotic cells were incubated in 10 ml hypotonic buffer (75 mM KCl, 0.2 mM spermine, 0.5 mM spermidine) for 10 minutes at room temperature to swell the cells. The cells were then centrifuged at 100 g for 8 minutes. The cell pellet was drained carefully and about 10^7 cells were resuspended in 1 ml polyamine buffer [15 mM Tris-HCl, 20 mM NaCl, 80 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 14 mM β -mercaptoethanol, 0.1% digitonin, 0.2 mM Spermine, 0.5 mM spermidine]. Chromosomes were then released by gently drawing the cell suspension up and expelling it through a 22 G needle attached to a 3 ml plastic syringe. The chromosome concentration was about $1-3 \times 10^8$ chromosomes/ml.

2. Hexylene glycol buffer system

In the second procedure, about 8×10^6 mitotic cells were resuspended in 10 ml glycine-hexylene glycol buffer [100 mM glycine, 1% hexylene glycol, pH 8.4-8.6 adjusted with saturated Ca-hydroxide solution] and incubated for 10 minutes at 37°C, followed by centrifugation for 10 minutes to pellet the nuclei. The supernatant was

C. Staining of chromosomes with DNA specific dyes

- #### D. Flow sorting of chromosomes

- #### E. Storage of the Sorted Artificial Chromosomes

- ## F. Quality Control

- ## 1. Analysis of the purity

-80-

DNA probe [see, Hadlaczky et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:8106-8110. Purity of the sorted chromosomes was 97–99%.

2. Characteristics of the sorted chromosomes

5 Pulsed field gel electrophoresis and Southern hybridization were carried out to determine the size distribution of the DNA content of the sorted artificial chromosomes.

C. Functioning of the artificial chromosomes

To check whether their activity is preserved, the artificial chromosomes are microinjected into primary cells, somatic cells and stem cells.

cells.

·II. **Sorting of mammalian artificial chromosome containing microcells**

A. Micronucleation

A. Micronucleation
Cells were grown to 80—90% confluency in 4 T150 flasks. Colcemid was added to a final concentration of 0.06 $\mu\text{g/ml}$, and then incubated with the cells at 37° C for 24 hours.

B. Enucleation

B. Nucleation
Ten $\mu\text{g/ml}$ cytochalasin B was added and the resulting microcells were centrifuged at 15,000 rpm for 70 minutes at 28–33° C.

C. Purification of microcells by filtration

20 The microcells were purified using Swinnex filter units and Nucleopore filters [$5\mu\text{m}$ and $3\mu\text{m}$].

D. Staining and sorting microcells

As above, the cells were stained Hoechst and chromomycin A3 dyes. The microcells were sorted by cell sorter to isolate the microcells that contain the mammalian artificial chromosome.

E. Fusion

E. Fusion

The microcells that contain the artificial chromosome are fused to selected primary cells, somatic cells, embryonic stem cells to generate transgenic animals for gene therapy purposes, and other cells to deliver the chromosomes to the cells.

EXAMPLE 11

Introduction of mammalian artificial chromosomes into insect cells

Insect cells should be useful hosts for MACs, particularly for production of gene products for a number of reasons, including:

- 5 1. A mammalian artificial chromosome provides extra genomic specific integration site for introduction of genes encoding proteins of interest [reduced chance of mutation in production system].
2. The large size of artificial chromosome permits megabase size DNA integration so that genes encoding an entire pathway leading to a
- 10 protein or nonprotein of therapeutic value, such as an alkaloid [digitalis, morphine, taxol].
3. Amplification of genes encoding useful proteins can be accomplished in the artificial mammalian chromosome to obtain higher protein yields in insect cells.
- 15 4. Insect cells support required post translational modifications (glycosylation, phosphorylation) essential for protein biological function.
5. Insect cells do not support mammalian viruses — eliminates cross-contamination of product with human infectious agents.
6. The ability to introduce chromosomes, circumvents
- 20 traditional recombinant baculovirus systems for production of nutritional, industrial or medicinal proteins in insect cell systems.
7. The low temperature optimum for insect cell growth (28° C) permits reduced energy cost of production.
8. Serum free growth medium for insect cells will result in
- 25 lower production costs.
9. Artificial chromosome containing cells can be stored indefinitely at low temperature.
10. Insect larvae will serve as biological factories for the production of nutritional, medicinal or industrial proteins by microinjection
- 30 of fertilized insect eggs.

A. Demonstration that insect cells recognize mammalian promoters

Gene constructs containing a mammalian promoter, such as CMV linked to DNA encoding a detectable marker gene fusions [*Renilla* luciferase gene (see, e.g., U.S. Patent No. 5,292,658 for a description of

- 5 DNA encoding the *Renilla* luciferase, and plasmid pTZrLuc-1, which can provide the starting material for construction of such vectors, see SEQ ID No. 10] and also including the simian virus 40 (SV40) promoter operably linked to the beta galactosidase gene] was introduced into the cells of two species *Trichoplusia ni* [cabbage looper] and *Bombyx mori* [silk worm].

- 10 After transferring the constructs into the insect cell lines either by electroporation or by microinjection, expression of the marker genes was detected after a 24 h incubation. In each case a positive result was obtained in the samples containing the genes which was absent in
- 15 samples in which the genes were omitted. In addition, a β -actin promoter - *Renilla* luciferase fusion was introduced into the *T. ni* and *B. mori* cells yielding light emission. Thus, mammalian promoters function to direct expression of these marker genes in insects. Therefore, MACs are candidates for expression of heterologous genes in insect cells.

20 **B. Construction of vectors for use in insect cells and fusion with mammalian cells**

1. Transform LMTK⁻ cells with expression vector with:
- a. *B. mori* β -actin promoter — Hyg^r selectable marker gene for insect cells, and.
- 25 b. SV40 or CMV promoters controlling a puromycin^r selectable marker gene for mammalian cells.
2. Detect expression of the mammalian promoter in LMTk cells (puromycin^r LMTk cells)
3. Use pur^r cells in fusion experiments with *Bombyx* and
- 30 *Trichoplusia* cells, select Hyg^r cells.

C. Insertion of the MACs into insect cells

These experiments are designed to detect expression of a detectable marker gene [such as β -gal expressed under the control of a mammalian promoter, such as pSV40] located on a MAC. Data indicate

5 that β -gal was expressed.

Insect cells of each species are fused with mammalian cells containing either the mini chromosome [EC3/7C5] or the mini and the megachromosome [such as GHB-42, which is a cell line recloned from G3D5] or a cell line that carries only the megachromosome [such as

10 H1D3 or a reclone therefrom]. Fusion is carried out as follows:

1. mammalian + insect cells (50/50%) in log phase growth are mixed;
2. calcium / PEG cell fusion: (10 min — 0.5 h);
3. heterokaryons (+ 72 h) are selected.

15 The following selection conditions to select for insect cells that contain a MAC can be used: [+ = positive selection; - = negative selection]:

1. growth at 28° C (+ insect cells, - mammalian cells);
2. Graces insect cell medium [SIGMA] (- mammalian cells);
- 20 3. no exogenous CO₂ (- mammalian cells); and/or
4. antibiotic selection (Hyg or G418) (+ transformed insect cells).

Immediately following the fusion protocol, many heterokaryons [fusion events] are observed between the mammalian and each species of insect cells [up to 90% heterokaryons]. After growth [2 + weeks] on

25 insect medium containing G418 and/or hygromycin at selection levels used for selection of transformed mammalian cells, individual colonies are detected growing on the fusion plates. By virtue of selection for the

antibiotic resistance conferred by the MAC and selection for insect cells, these colonies should contain MACs.

EXAMPLE 12

5 Preparation of the chromosome fragmentation vector and other vectors for targeted integration of DNA into MACs

Fragmentation of the megachromosome, should ultimately result in smaller stable chromosomes that contain about 15 Mb to 50 Mb that will be easily manipulated for use as vectors. Vectors to effect such fragmentation should also aid in determination and identification of the
10 elements required for preparation of an in vitro-produced artificial chromosome.

Reduction in the size of the megachromosome can be achieved in a number of different ways including: stress treatment, such as by starvation, or cold or heat treatment; treatment with agents that
15 destabilize the genome or nick DNA, such as BrdU, coumarin, EMS and others; treatment with ionizing radiation [see, e.g., Brown (1992) Curr. Opin. Genes Dev. 2:479-486]; and telomere-directed in vivo chromosome fragmentation [see, e.g., Far et al. (1995) EMBO J. 14:5444-5454].

20 A. Preparation vectors for fragmentation of the artificial chromosome and also for targeted integration of selected gene products

1. Construction of pTEMPUD

Plasmid pTEMPUD [see, Figure 5] is a mouse homologous recombination "killer" vector for in vivo chromosome fragmentation, and
25 also for inducing large-scale amplification via site specific integration. With reference to Figure 5, the PstI to Sall fragment was derived from pBabe-puro retroviral vector [see, Morgenstern et al. (1990) Nucleic Acids Res. 18:3587-3596]. This fragment contains DNA encoding ampicillin resistance, the pUC origin of replication, and the puromycin N-acetyl
30 transferase gene under control of the SV40 early promoter. The URA3 portion comes from the pYAC5 cloning vector [SIGMA]. URA3 was cut

out of pYAC5 with Sall-XhoI digestion, cloned into pNEB193 [New England Biolabs], which was then cut with EcoRI-Sall and ligated to the Sall site of pBabepuro to produce pPU.

- 5 A 1293 bp fragment [see SEQ ID No. 1] encoding the mouse major satellite, was isolated as an EcoRI fragment from a DNA library produced from mouse LMTK fibroblast cells and inserted into the EcoRI site of pPU to produce pMPU.

- 10 The TK promoter driven diphtheria toxin gene [DT-A] was derived from pMC1DT-A [see, Maxwell *et al.* (1986) *Cancer Res.* 46:4660-4666] by BglII-XhoI digestion and cloned into the pMC1neo poly A expression vector [STRATAGENE, La Jolla, CA] by replacing the neo coding sequence. The TK promoter, DT-A gene and poly A sequence were removed from this vector, cohesive ends were filled with Klenow and the resulting fragment blunt end-ligated and ligated into the SnaBI [TACGTA]
- 15 of pMPU to produce pMPUD.

- The Hutel 2.5 fragment [see SEQ ID No.3] was inserted at the PstI site [see the 6100 PstI - 3625 PstI fragment on pTEMPUD] of pMPUD to produce pTEMPUD. This fragment includes a human telomere. It includes a unique BglII site [see nucleotides 1042-1047 of SEQ ID No.3],
- 20 which will be used as a site for introduction of a synthetic telomere that will include multiple repeats [80] of GGGATT with BamHI and BglII ends for insertion into the BglII site which will then remain unique, since the BamHI overhang is compatible with the BglII site. Ligation of BamHI fragment to a BglII destroys the BglII site, so that only a single BglII site
- 25 will remain. Selection for the unique BglII site insures that the synthetic telomere will be inserted in the correct orientation. The unique BglII site is the site at which the vector is linearized.

2. Use of pTEMPUD for in vivo chromosome fragmentation

Linearization of pTEMPUD by BglII results in a linear molecule with a human telomere at one end. Integration of this linear fragment into the chromosome, such as the megachromosome in hybrid cells or any mouse chromosome, which contains repeats of the mouse major satellite sequence, results in integration of the selectable marker puromycin and cleavage of the plasmid by virtue of the telomeric end. The DT gene prevents that entire linear fragment from integrating by random events, since upon integration and expression it is toxic. Thus random integration will be toxic. Thus, site directed integration into the targeted DNA will be selected. Such integration will produce fragmented chromosomes.

The fragmented truncated chromosome with the new telomere will survive, and the other fragment without the centromere will be lost. Repeated in vivo fragmentations will ultimately result in selection of the smallest functioning minichromosome possible.

Thus this vector can be used to produce minichromosomes from mouse chromosomes, or to fragment the megachromosome.

3. pTEMPhu and pTEMPhu3

Vectors that specifically target human chromosomes can be constructed from pTEMPUD. These vectors can be used to fragment specific human chromosomes, depending upon the selected satellite sequence, to produce human minichromosomes, and also to isolate human centromeres.

a. pTEMPhu

To render pTEMPUD suitable for fragmenting human chromosomes, the mouse major satellite sequence is replaced with human satellite sequences. Unlike mouse chromosomes, each human chromosome has a unique satellite sequence. For example, the mouse major satellite has been replaced with a human hexameric α -satellite [or alphoid satellite] DNA sequence. This sequence is an 813 bp fragment [nucleotide 232-1044 of

SEQ ID No. 2] from clone pS12, deposited in the EMBL database under Accession number X60716, isolated from a human colon carcinoma cell line Colo320 [deposited under Accession No. ATCC CCL 220.1]. The 813 bp alphoid fragment can be obtained from the pS12 clone by nucleic acid amplification using synthetic primers, which each contain an EcoRI site, as follows:

GGGGAATTCAT TGGGATGTTT CAGTTGA forward primer [SEQ ID No. 4]
CGAAAGTCCCC CTTAGGAGAT CTTAAGGA reverse primer [SEQ ID No. 5].

Digestion of the amplified product with EcoRI results in a fragment with EcoRI ends that includes the human α -satellite sequence. This sequence is inserted into pTEMPUD in place of the EcoRI fragment that contains the mouse major satellite.

b. pTEMPHu3

In pTEMPHu3, the mouse major satellite sequence is replaced by the 3kb human chromosome 3-specific α -satellite from D3Z1 [deposited under ATCC Accession No. 85434; see, also Yrokov (1989) Cytogenet. Cell Genet. 51:1114].

4. Use of the pTEMPHU3 to induce amplification on human chromosome #3

Each human chromosome contains unique chromosome-specific alphoid sequence. Thus, use of pTEMPHU3, which is targeted to the chromosome 3-specific α -satellite can be introduced into human cells under selective conditions, whereby large scale amplification of the chromosome 3 centromeric region and production of a *de novo* chromosome. Such induced large-scale amplification provides a means for inducing *de novo* chromosome formation and also for *in vivo* cloning of defined human chromosome fragments up to megabase size.

For example, the break-point in human chromosome #3 is on the short arm near the centromere. this region is involved in renal cell carcinoma formation. By targeting pTEMPHu3 to this region, the induced

large-scale amplification may contain this region, which can then be cloned using the bacterial and yeast markers in the pTEMPhu3 vector.

- The pTEMPhu3 cloning vector allows not only selection for homologous recombinants, but also direct cloning of the integration site in
- 5 YACS. This vector can also be used to target human chromosome #3, preferably with a deleted short arm, in a mouse-human monochromosomal microcell hybrid line. Homologous recombinants can be screened by nucleic acid amplification (PCR) and amplification can be screened by DNA hybridization, Southern hybridization, and *in situ* hybridization. The
- 10 amplified region can be cloned into YAC. This vector and these methods also permit a functional analysis of cloned chromosome regions by reintroducing the cloned amplified region into mammalian cells.

B. Preparation of libraries in YAC vectors for cloning of centromeres and identification of functional chromosomal units

- 15 Another method that may be used to obtain smaller-sized functional mammalian artificial chromosome units and to clone centromeric DNA involves screening of mammalian DNA YAC vector-based libraries and functional analysis of potential positive clones in a transgenic mouse model system. A mammalian DNA library is prepared in a YAC vector,
- 20 such as YRT2 [see Schedl *et al.* (1993) Nuc. Acids Res. 21:4783-4787], which contains the murine tyrosinase gene. The library is screened for hybridization to mammalian telomere and centromere sequence probes. Positive clones are isolated and microinjected into pronuclei of fertilized oocytes of NMRI/Han mice following standard techniques. The embryos
- 25 are then transferred into NMRI/Han foster mothers. Expression of the tyrosinase gene in transgenic offspring confers an identifiable phenotype (pigmentation). The clones that give rise to tyrosinase-expressing transgenic mice are thus confirmed as containing functional mammalian artificial chromosome units.

- 30 Alternatively, fragments of SATACs may be introduced into the YAC vectors and then introduced into pronuclei of fertilized oocytes of

NMRI/Han mice following standard techniques as above. The clones that give rise to tyrosinase-expressing transgenic mice are thus confirmed as containing functional mammalian artificial chromosome units, particularly centromeres.

5 C. **Incorporation of Heterologous Genes into Mammalian Artificial Chromosomes through The Use of Homology Targeting Vectors**

As described above, the use of mammalian artificial chromosomes for expression of heterologous genes obviates certain negative effects that may result from random integration of heterologous plasmid DNA into the recipient cell genome. An essential feature of the mammalian artificial chromosome that makes it a useful tool in avoiding the negative effects of random integration is its presence as an extra-genomic gene source in recipient cells. Accordingly, methods of specific, targeted incorporation of heterologous genes exclusively into the mammalian artificial chromosome, without extraneous random integration into the genome of recipient cells, are desired for heterologous gene expression from a mammalian artificial chromosome.

One means of achieving site-specific integration of heterologous genes into artificial chromosomes is through the use of homology targeting vectors. The heterologous gene of interest is subcloned into a targeting vector which contains nucleic acid sequences that are homologous to nucleotides present in the artificial chromosome. The vector is then introduced into cells containing the artificial chromosome for specific site-directed integration into the artificial chromosome through a recombination event at sites of homology between the vector and the chromosome. The homology targeting vectors may also contain selectable markers for ease of identifying cells that have incorporated the vector into the artificial chromosome as well as lethal selection genes that are expressed only upon extraneous integration of vector into the recipient cell genome. Two exemplary homology targeting vectors, λ CF-7 and p λ CF-7-DTA, are described below.

1. Construction of Vector λ CF-7

Vector λ CF-7 contains the cystic fibrosis transmembrane conductance regulator [CFTR] gene as an exemplary therapeutic molecule-encoding nucleic acid that may be incorporated into mammalian artificial chromosomes for use in gene therapy applications. This vector, which also contains the puromycin-resistance gene as a selectable marker, as well as the *Saccharomyces cerevisiae* *ura3* gene [orotidine-5-phosphate decarboxylase], was constructed in a series of steps as follows.

a. Construction of pURA

10 Plasmid pURA was prepared by ligating a 2.6-kb Sall/XhoI fragment from the yeast artificial chromosome vector pYAC5 [Sigma; see also Burke *et al.* (1987) *Science* 236:806-812 for a description of YAC vectors as well as GenBank Accession no. U01086 for the complete sequence of pYAC5] containing the *S. cerevisiae* *ura3* gene with a 3.3-kb 15 Sall/SmaI fragment of pHyg [see, *e.g.*, U.S. Patent Nos. 4,997,764, 4,686,186 and 5,162,215, and the description above]. Prior to ligation the XhoI end was treated with Klenow polymerase for blunt end ligation to the SmaI end of the 3.3 kb fragment of pHyg. Thus, pURA contains the *S. cerevisiae* *ura3* gene, and the *E. coli* ColE1 origin of replication and 20 the ampicillin-resistance gene. The *uraE* gene is included to provide a means to recover the integrated construct from a mammalian cell as a YAC clone.

b. Construction of pUP2

Plasmid pURA was digested with Sall and ligated to a 1.5-kb 25 Sall fragment of pCEPUR. Plasmid pCEPUR is produced by ligating the 1.1 kb SnaBI-NhaI fragment of pBabe-puro [Morgenstern *et al.* (1990) *Nucl. Acids Res.* 18:3587-3596; provided by Dr. L. Székely (Microbiology and Tumorbiology Center, Karolinska Institutet, Stockholm); see, also Tonghua *et al.* (1995) *Chin. Med. J.* (Beijing, Engl. Ed.) 108:653-659; 30 Couto *et al.* (1994) *Infect. Immun.* 62:2375-2378; Dunckley *et al.* (1992)

FEBS Lett. 296:128-34; French *et al.* (1995) *Anal. Biochem.* 228:354-355; Liu *et al.* (1995) *Blood* 85:1095-1103; International PCT application Nos. WO 9520044; WO 9500178, and WO 9419456] to the *NheI-NruI* fragment of pCEP4 [Invitrogen].

- 5 The resulting plasmid, pUP2, contains the all the elements of pURA plus the puromycin-resistance gene linked to the SV40 promoter and polyadenylation signal from pCEPUR.

c. **Construction of pUP-CFTR**

- The intermediate plasmid pUP-CFTR was generated in order to combine the elements of pUP2 into a plasmid along with the CFTR gene. First, a 4.5-kb *Sall* fragment of pCMV-CFTR that contains the CFTR-encoding DNA [see, also, Riordan *et al.* (1989) *Science* 245:1066-1073, U.S. Patent No. 5,240,846, and Genbank Accession no. M28668 for the sequence of the CFTR gene] containing the CFTR gene only was
- 15 ligated to *XhoI*-digested pCEP4 [Invitrogen and also described herein] in order to insert the CFTR gene in the multiple cloning site of the Epstein Barr virus-based (EBV) vector pCEP4 [Invitrogen, San Diego, CA; see also Yates *et al.* (1985) *Nature* 313:812-815; see, also U.S. Patent No. 5,468,615] between the CMV promoter and SV40 polyadenylation signal.
- 20 The resulting plasmid was designated pCEP-CFTR. Plasmid pCEP-CFTR was then digested with *Sall* and the 5.8-kb fragment containing the CFTR gene flanked by the CMV promoter and SV40 polyadenylation signal was ligated to *Sall*-digested pUP2 to generate pUP-CFTR. Thus, pUP-CFTR contains all elements of pUP2 plus the CFTR gene linked to the CMV
- 25 promoter and SV40 polyadenylation signal.

d. **Construction of Δ CF-7**

- Plasmid pUP-CFTR was then linearized by partial digestion with *EcoRI* and the 13 kb fragment containing the CFTR gene was ligated with *EcoRI*-digested Charon 4A λ [see Blattner *et al.* (1977) *Science*
- 30 196:161; Williams and Blattner (1979) *J. Virol.* 29:555 and Sambrook *et*

- al. (1989) Molecular Cloning, A Laboratory Manual, Second Ed., Cold Spring Harbor Laboratory Press, Volume 1, Section 2.18, for descriptions of Charon 4A λ]. The resulting vector, λ CF8, contains the Charon 4A λ bacteriophage left arm, the CFTR gene linked to the CMV promoter and
- 5 SV40 polyadenylation signal, the ura3 gene, the puromycin-resistance gene linked to the SV40 promoter and polyadenylation signal, the thymidine kinase promoter [TK], the ColE1 origin of replicaton, the ampicillan resistance gene and the Charon 4A λ bacteriophage right arm. The λ CF8 construct was then digested with XhoI and the resulting 27.1
- 10 kb was ligated to the 0.4kb XhoI/EcoRI fragment of pJBP86 [described below], containing the SV40 polyA signal and the EcoRI-digested Charon 4A λ right arm. The resulting vector λ CF-7 contains the Charon 4A λ left arm, the CFTR encoding DNA linked to the CMV promoter and SV40 polyA signal, the ura3 gene, the puromycin resistance gene linked to the
- 15 SV40 promoter and polyA signal adn the Charon 4A λ right arm. The λ DNA fragments provide sequences homologous to nucleotides present in the exemplary artificial chromosomes.

- The vector is then introduced into cells containing the artificial chromosomes exemplified herein. Accordingly, when the linear λ CF-7
- 20 vector is introduced into megachromosome-carrying fusion cell lines, such as described herein, it will be specifically integrated into the megachromosome through recombination between the homologous bacteriophage λ sequences of the vector and the artificial chromosome.

2. Construction of Vector λ CF-7-DTA

- 25 Vector λ CF-7-DTA also contains all the elements contained in λ CF-7, but additionally contains a lethal selection marker, the diptheria toxin-A (DT-A) gene as well as the ampicillin-resistance gene and an origin of replication. This vector was constructed in a series of steps as follows.

a. Construction of pJBP86

Plasmid pJBP86 was used in the construction of λ CF-7, above. A 1.5-kb Sall fragment of pCEPUR containing the puromycin-resistance gene linked to the SV40 promoter and polyadenylation signal was ligated to

5 HindIII-digested pJB8 [see, e.g., Ish-Horowitz *et al.* (1981) Nucleic Acids Res. 9:2989-2998; available from ATCC as Accession No. 37074; commercially available from Amersham, Arlington Heights, IL]. Prior to ligation the Sall ends of the 1.5 kb fragment of pCEPUR and the HindIII

10 linearized pJB8 ends were treated with Klenow polymerase before ligation. The resulting vector pJBP86 contains the puromycin resistance gene linked to the SV40 promoter and polyA signal, the 1.8 kb COS region of Charon 4A λ , the ColE1 origin of replication and the ampicillin resistance gene.

b. Construction of pMEP-DTA

15 A 1.1-kb XhoI/Sall fragment of pMC1-DT-A [see, e.g., Maxwell *et al.* (1986) Cancer Res. 46:4660-4666] containing the diphtheria toxin-A gene was ligated to XhoI-digested pMEP4 [Invitrogen, San Diego, CA] to generate pMEP-DTA. To produce pMC1-DT-A, the coding region of the DTA gene was isolated as a 800 bp PstI/HindIII fragment from p2249-1

20 and inserted into pMC1neopolyA [pMC1 available from Stratagene] in place of the neo gene and under the control of the TK promoter. The resulting construct pMC1DT-A was digested with HindIII, the ends filled by Klenow and Sall linkers were ligated to produce a 1061 bp TK-DTA gene cassette with an XhoI end [5'] and a Sall end containing the 270 bp

25 TK promoter and the ~790 bp DT-A fragment. This fragment was ligated into XhoI-digested pMEP4.

Plasmid pMEP-DTA thus contains the DT-A gene linked to the TK promoter and SV40, ColE1 origin of replication and the ampicillin-resistance gene.

c. Construction of pJB83-DTA9

Plasmid pJB8 was digested with HindIII and Clal and ligated with an oligonucleotide [see SEQ ID NOs. 7 and 8 for the sense and antisense strands of the oligonucleotide, respectively] to generate pJB83.

- 5 The oligonucleotide that was ligated to Clal/HindIII-digested pJB8 contained the recognition sites of SwaI, PacI and SrfI restriction endonucleases. These sites will permit ready linearization of the pACF-7-DTA construct.

Next, a 1.4-kb XhoI/SalI fragment of pMEP-DTA, containing the

- 10 DT-A gene was ligated to Sall-digested pJB83 to generate pJB83-DTA9.

d. Construction of λ CF-7-DTA

The 12-bp overhangs of λ CF-7 were removed by Mung bean nuclease and subsequent T4 polymerase treatments. The resulting 41.1-kb linear λ CF-7 vector was then ligated to pFB83-DTA9 which had been

- 15 kb linear λ CF-7 vector was then ligated to pFB83-DTA9 which had been digested with ClaI and treated with T4 polymerase. The resulting vector, λ CF-7-DTA, contains all the elements of λ CF-7 as well as the DT-A gene linked to the TK promoter and the SV40 polyadenylation signal, the 1.8 kb Charon 4A λ COS region, the ampicillin-resistance gene [from pJB83-DTA9] and the Col E1 origin of replication [from pJB83-DT9A].

20 3. pMCT-RUC,

Plasmid pMCT-RUC [14kbp] was constructed for site-specific targeting of the Renilla luciferase [see, e.g., U.S. Patent Nos. 5,292,658 and 5,418,155 for a description of DNA encoding *Renilla* luciferase, and plasmid pTZrLuc-1, which can provide the starting material for

- 25 construction of such vectors] gene to a mammalian chromosome. The relevant features of this plasmid are the Renilla luciferase gene under transcriptional control of the human cytomegalovirus immediate-early gene enhancer/promoter; the hygromycin gene under the transcriptional control of the thymidine kinase promoter; and a unique HpaI site is for
30 linearizing the plasmid.

- This construct was introduced into C5 cells [see, Lorenz *et al.* (1996) *J. Biolum. Chemilum.* 11:31-37]. C5 mouse fibroblasts were maintained as a monolayer [see, Gluzman (1981) *Cell* 23:175-183]. Cells at 50% confluency in 100 mm Petri dishes were used for calcium phosphate transfection [see, Harper *et al.* (1981) *Chromosoma* 83:431-439] using 10 μ g of linearized pMCT-RUC per plate. Colonies originating from single transfected cells were isolated and maintained in F-12 medium containing hygromycin (300 μ g/mL) and 10% fetal bovine serum. Cells were grown in 100 mm Petri dishes prior to the Renilla luciferase assay.
- 5 The Renilla luciferase assay was performed [see, *e.g.*, Matthews *et al.* (1977) *Biochemistry* 16:85-91]. Hygromycin-resistant cell lines obtained after transfection of mouse fibroblasts with linearized plasmid pMCT-RUC ["B" cell lines] were grown to 100% confluency for measurements of light emission *in vivo* and *in vitro*. Light emission was measured
- 10 *in vivo* after about 30 generations as follows: growth medium was removed and replaced by 1 mL RPMI 1640 containing coelenterazine [1 mmol/L final concentration]. Light emission from cells was then visualized by placing the Petri dishes in a low light video image analyzer [Hamamatsu Argus-100]. An image was formed after 5 min. of photon
- 20 accumulation using 100% sensitivity of the photon counting tube. For measuring light emission *in vitro*, cells were trypsinized and harvested from one Petri dish, pelleted, resuspended in 1 mL assay buffer [0.5 mol/L NaCl, 1 mmol/L EDTA, 0.1 mol/L potassium phosphate, pH 7.4] and sonicated on ice for 10 s. Lysates were then assayed in a Turner TD-20e
- 25 lukminometer for 10 s after rapid injection of 0.5 mL of 1 mmol/L coelenterazine, and the average value of light emission was recorded as LU [1 LU = 1.6×10^6 hu/s for this instrument].

Independent cell lines of mouse fibroblasts transfected with linearized plasmid pMCT-RUC showed different levels of Renilla luciferase

30 activity. Similar differences in light emission were observed when

measurements were performed on lysates of the same cell lines. This variation in light emission was probably due to a position effect resulting from the random integration of plasmid pMCT-RUC into the mouse genome, since enrichment for site targeting of the luciferase gene was not performed in this experiment.

D. Protein secretion targeting vectors

Isolation of heterologous proteins produced intracellularly in mammalian cell expression systems requires cell disruption under potentially harsh conditions and purification of the recombinant protein from cellular contaminants. The process of protein isolation may be greatly facilitated by secretion of the recombinantly produced protein into the extracellular medium where there are fewer contaminants to remove during purification. Therefore, secretion targeting vectors have been constructed for use with the mammalian artificial chromosome system.

A useful model vector for demonstrating production and secretion of heterologous protein in mammalian cells contains DNA encoding a readily detectable reporter protein fused to an efficient secretion signal that directs transport of the protein to the cell membrane and secretion of the protein from the cell. Vectors pLNCX-ILRUC and pLNCX-ILRUC1, described below, are examples of such vectors. These vectors contain DNA encoding an interleukin-2 (IL2) signal peptide-Renilla reniformis luciferase fusion protein. The IL-2 signal peptide [encoded by the sequence set forth in SEQ ID No. 9] directs secretion of the luciferase protein, to which it is linked, from mammalian cells. Upon secretion from the host mammalian cell, the IL-2 signal peptide is cleaved from the fusion protein to deliver mature, active, luciferase protein to the extracellular medium. Successful production and secretion of this heterologous protein can be readily detected by performing luciferase assays which measure the light emitted upon exposure of the medium to the bioluminescent luciferin substrate of the luciferase enzyme.

1. Construction of Protein Secretion Vector pLNCX-ILRUC

Vector pLNCX-ILRUC contains a human IL-2 signal peptide-R. reniformis fusion gene linked to the human cytomegalovirus (CMV) immediate early promoter for constitutive expression of the gene in mammalian cells. The construct was prepared as follows.

a. Preparation of the IL-2 signal sequence-encoding DNA

A 69-bp DNA fragment containing DNA encoding the human IL-2 signal peptide was obtained through nucleic acid amplification, using appropriate primers for IL-2, of an HEK 293 cell line [see, e.g., U.S. Patent No. 4,518,584 for an IL-2 encoding DNA; see, also SEQ ID No. 9; the IL-2 gene and corresponding amino acid sequence is also provided in the Genbank Sequence Database as accession nos. K02056 and J00264]. The signal peptide includes the first 20 amino acids shown in the translations provided in both of these Genbank entries and in SEQ ID NO. 9. The corresponding nucleotide sequence encoding the first 20 amino acids is also provided in these entries [see, e.g., nucleotides 293-52 of accession no. K02056 and nucleotides 478-537 of accession no. J00264], as well as in SEQ ID NO. 9. The amplification primers included an EcoRI site [GAATTC] that for subcloning of the DNA fragment into EcoRI-digested pGEMT [Promega]. The forward primer is set forth in SEQ ID No. 11 and the sequence of the reverse primer is set forth in SEQ ID No. 12.

TTTGAATTCATGTACAGGATGCAACTCCTG forward [SEQ ID No. 11]

TTTGAATTCAGTAGGTGCACTGTTTGTGAC reverse [SEQ ID No. 12]

b. Preparation of the R. reniformis luciferase-encoding DNA

The initial source of the R. reniformis luciferase gene was plasmid pLXSN-RUC. Vector pLXSN [see, e.g., U.S. Patent Nos. 5,324,655, 5,470,730, 5,468,634, 5,358,866 and Miller et al. (1989) Biotechniques 7:980] is a retroviral vector capable of expressing heterologous DNA under the transcriptional control of the retroviral LTR; it

also contains the neomycin-resistance gene operatively linked for expression to the SV40 early region promoter. The *R. reniformis* luciferase gene was obtained from plasmid pTZrLuc-1 [see, e.g., U.S. Patent No. 5,292,658; see also the Genbank Sequence Database accession no. M63501; and see also Lorenz et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:4438-4442] and is shown as SEQ ID NO. 10. The 0.97 kb EcoRI/SmaI fragment of pTZrLuc-1 contains the coding region of the *Renilla* luciferase-encoding DNA. Vector pLXSN was digested with

10 contains the luciferase gene located operably linked to the viral LTR and upstream of the SV40 promoter, which directs expression of the neomycin-resistance gene.

c. **Fusion of DNA encoding the IL-2 Signal Peptide and the *R. reniformis* Luciferase Gene to Yield pLXSN-ILRUC**

15 The pGEMT vector containing the IL-2 signal peptide-encoding DNA described in 1.a. above was digested with EcoRI, and the resulting fragment encoding the signal peptide was ligated to EcoRI-digested pLXSN-RUC. The resulting plasmid, called pLXSN-ILRUC, contains the IL-

20 2 signal peptide-encoding DNA located immediately upstream of the *R. reniformis* gene in pLXSN-RUC. Plasmid pLXSN-ILRUC was then used as a template for nucleic acid amplification of the fusion gene in order to add a SmaI site at the 3' end of the fusion gene. The amplification product was subcloned into EcoRI/SmaI-digested pGEMT (Promega) to generate

25 ILRUC-pGEMT.

d. **Introduction of the Fusion Gene into a Vector Containing Control Elements for Expression in Mammalian Cells**

Plasmid ILRUC-pGEMT was digested with KspI and SmaI to

30 release a fragment containing the IL-2 signal peptide-luciferase fusion gene which was ligated to HpaI-digested pLNCX. Vector pLNCX [see, e.g., U.S. Patent Nos. 5,324,655 and 5,457,182; see, also Miller and

Rosman (1989) Biotechniques 7:980-990] is a retroviral vector for expressing heterologous DNA under the control of the CMV promoter; it also contains the neomycin-resistance gene under the transcriptional control of a viral promoter. The vector resulting from the ligation reaction was designated pLNCX-ILRUC. Vector pLNCX-ILRUC contains the IL-2 signal peptide-luciferase fusion gene located immediately downstream of the CMV promoter and upstream of the viral 3' LTR and polyadenylation signal in pLNCX. This arrangement provides for expression of the fusion gene under the control of the CMV promoter. Placement of the heterologous protein-encoding DNA [i.e., the luciferase gene] in operative linkage with the IL-2 signal peptide-encoding DNA provides for expression of the fusion in mammalian cells transfected with the vector such that the heterologous protein is secreted from the host cell into the extracellular medium.

15 2. **Construction of Protein Secretion Targeting Vector pLNCX-ILRUC**

Vector pLNCX-ILRUC may be modified so that it can be used to introduce IL-2 signal peptide-luciferase fusion gene into a mammalian artificial chromosome in a host cell. To facilitate specific incorporation of the pLNCX-ILRUC expression vector into a mammalian artificial chromosome, nucleic acid sequences that are homologous to nucleotides present in the artificial chromosome are added to the vector to permit site directed recombination.

Exemplary artificial chromosomes described herein contain lambda phage DNA. Therefore, protein secretion targeting vector pLNCX-ILRUC was prepared by addition of lambda phage DNA [from Charon 4A arms] to produce the secretion vector pLNCX-ILRUC.

3. Expression and Secretion of R. reniformis Luciferase from Mammalian Cells

a. Expression of R. reniformis Luciferase Using pLNCX-ILRUC

5 Mammalian cells [LMTK⁻ from the ATCC] were transiently transfected with vector pLNCX-ILRUC [$\sim 10 \mu\text{g}$] by electroporation [BIORAD, performed according to the manufacturer's instructions]. Stable transfectants produced by growth in G418 for neo selection have also been prepared.

10 Transfectants were grown and then analyzed for expression of luciferase. To determine whether active luciferase was secreted from the transfected cells, culture media were assayed for luciferase by addition of coelentraine [see, e.g., Matthews *et al.* (1977) *Biochemistry* 16:85-91].

The results of these assays establish that vector pLNCX-ILRUC is 15 capable of providing constitutive expression of heterologous DNA in mammalian host cells. Furthermore, the results demonstrate that the human IL-2 signal peptide is capable of directing secretion of proteins fused to the C-terminus of the peptide. Additionally, these data demonstrate that the R. reniformis luciferase protein is a highly effective 20 reporter molecule, which is stable in a mammalian cell environment, and forms the basis of a sensitive, facile assay for gene expression.

b. Expression of R. reniformis Luciferase Using pLNCX-ILRUC λ

To express the IL-2 signal peptide-R. reniformis fusion gene from an 25 artificial mammalian chromosome, vector pLNCX-ILRUC λ is targeted for site-specific integration into an artificial mammalian chromosome through homologous recombination of the lambda DNA sequences contained in the chromosome and the vector. This is accomplished by introduction of pLNCX-ILRUC λ into either a fusion cell line harboring mammalian artificial 30 chromosomes or mammalian host cells that contain artificial mammalian chromosomes. If the vector is introduced into a fusion cell line harboring

the artificial chromosomes, for example through microinjection of the vector or transfection of the fusion cell line with the vector, the cells are then grown under selective conditions, i.e. artificial chromosomes, which have incorporated vector pLNCX-ILRUC λ , are isolated from the surviving cells, using purification procedures as described above, and then injected into the mammalian host cells.

Alternatively, the mammalian host cells may first be injected with mammalian artificial chromosomes which have been isolated from a fusion cell line. The host cells are then transfected with vector pLNCX-ILRUC λ and grown.

The recombinant host cells are then assayed for luciferase expression as described above.

D. Other targeting vectors

These vector rely on positive and negative selection to insure insertion and selection for the double recombinants. A single crossover results in incorporation of the DT-A, which kills the cell, double crossover recombinations delete the DT-1 gene.

1. Plasmid pNEM1 contains:
 - DT-A: Diphtheria toxin gene (negative selectable marker)
 - 20 Hyg: Hygromycin gene (positive selectable marker)
 - ruc: *Renilla* luciferase gene (non-selectable marker)
 - 1: LTR-MMTV promoter
 - 2: TK promoter
 - 3: CMV promoter
 - 25 MMR: Homology region (plasmid pAG60)
2. plasmid pNEM-2 and -3 are similar to pNEM 1 except for different negative selectable markers:
 - pNEM-1: diphtheria toxin gene as "—" selectable marker
 - pNEM-2: hygromycin antisense gene as "—" selectable marker
 - 30 pNEM-3: thymidine kinase HSV-1 gene as "—" selectable marker

3. Plasmid - lambda DNA based homology:

pNEM λ -1: base vector

pNEM λ -2: base vector containing p5 = gene

1: LTR MMTV promoter

2: SV40 promoter

3: CMV promoter

4: μ TIIA promoter (metallothionein gene promoter)

— homology region (plasmid pAG60)

λ L.A. and λ R.A. homology regions for λ left and right arms

10 (λ gt-WES).

EXAMPLE 13

Microinjection of mammalian cells with plasmid DNA

These procedure will be used to microinject MACS into eukaryotic cells, including mammalian and insect cells.

15 The microinjection technique is based on the use of small glass capillaries as a delivery system into cells and has been used for introduction of DNA fragments into nuclei [see, e.g., Chalfie et al. (1994) Science 263:802-804]. It allows the transfer of almost any type of molecules, e.g., hormones, proteins, DNA and RNA, into either the

20 cytoplasm or nuclei of recipient cells. This technique has no cell type restriction and is more efficient than other methods, including Ca^{2+} —mediated gene transfer and liposome-mediated gene transfer. About 20-30% of the injected cells become successfully transformed.

Microinjection is performed under a phase-contrast microscope. A

25 glass microcapillary, prefilled with the DNA sample, is directed into a cell to be injected with the aid of a micromanipulator. An appropriate sample volume (1-10 pl) is transferred into the cell by gentle air pressure exerted by a transjector connected to the capillary. Recipient cells are grown on glass slides imprinted with numbered squares for convenient localization

30 of the injected cells.

Nunclon tissue culture dishes 35 x 10 mm, Mouse cell line EC3/7C5 Plasmid DNA pCH110 [Pharmacia], Purified Green Florescent Protein (GFP) [GFPs from *Aequorea* and *Renilla* have been purified and also DNA encoding GFPs has been cloned; see, e.g., Prasher et al. (1992) Gene 111:229-233; International PCT Application No. WO 95/07463, which is based on U.S. application Serial No. 08/119,678 and U.S. application Serial No. 08/192,274], ZEISS Axiovert 100 microscope, Eppendorf transjector 5246, Eppendorf micromanipulator 5171, Eppendorf Cellocate coverslips, Eppendorf microloaders, Eppendorf femtotips and other standard equipment

b. Protocol.

- (1) Fibroblast cells are grown in Ø 35 mm tissue culture dishes (37° C, 5% CO₂) until the cell density reaches 80% confluency. The dishes are removed from the incubator and medium to added to about a 5 mm depth.
- (2) The dish is placed onto the dish holder and the cells observed with 10 x objective; the focus is desirably above the cell surface.
- (3) Plasmid or chromosomal DNA solution [1 ng/μl] and GFP protein solution are further purified by centrifuging the DNA sample at force sufficient to removed any particular debris [typically about 10,000 rpm for 10 minutes in a microcentrifuge].
- (4) Two 2 μl of the DNA solution (1 ng/μl) is loaded into a microcapillary with an Eppendorf microloader. During loading, the loader is inserted to the tip end of the microcapillary. GFP (1 mg/ml) is loaded wit the same procedure.
- (5) The protecting sheath is removed from the microcapillary and the microcapillary is fixed onto the capillary holer connected with the micromanipulator.

(6) The capillary tip lowered to the surface of the medium and focus on the cells gradually until the tip of the capillary reaches the surface of a cell. Lower the capillary further so that the capillary is inserted into the cell. Various parameters, such as the level of the capillary, the time and pressure are determined for the particular equipment. For example, using the fibroblast cell line C5 and the above-noted equipment, the best conditions are: injection time 0.4 second, pressure 80 psi. DNA can then be automatically injected into the nuclei of the cells.

(7) After injection, the cells are returned to the incubator, and incubated for about 18-24 hours.

(8) After incubation the number of transformants can be determined by a suitable method, which depends upon the selection marker. For example, if green fluorescent protein is used, the assay can be performed using UV light source and fluorescent filter set at 0-24 hours after injection. If β -gal-containing DNA, such as DNA-derived from pHCl10, has been injected, then the transformants can be assayed for β -gal.

i. Detection of β -galactosidase in cells injected with plasmid DNA. The medium is removed from the culture plate and the cells are fixed by addition of 5 ml of fixation Solution I: (1% glutaraldehyde; 0.1 M sodium phosphate buffer, pH 7.0; 1 mM MgCl_2), and incubated for 15 minutes at 37° C. Fixation Solution I is replaced with 5 ml of X-gal Solution II: [0.2% X-gal, 10 mM sodium phosphate buffer (pH 7.0), 150 mM NaCl, 1 mM MgCl_2 , 3.3 mM $\text{K}_4\text{Fe}(\text{CN})_6 \cdot \text{H}_2\text{O}$, 3.3 mM $\text{K}_3\text{Fe}(\text{CN})_6$], and the plates are incubated for 30-60 minutes at 37° C.

The X-gal solution is removed and 2 ml of 70% glycerol is added to each dish. Blue stained cells are identified under a light microscope.

This will be used to introduce a MAC, particular the MAC with the anti-HIV megachromosome, to produce a mouse model for anti-HIV activity.

EXAMPLE 14

5 Transgenic animals

Transgenic animals can be generated that express heterologous genes which confer desired traits, e.g., disease resistance, in the animals. A transgenic mouse is prepared to serve as model of a disease-resistant animal. Genes that encode vaccines or that encode therapeutic molecules
10 can be introduced into embryos or ES cells to produce animals that express the gene product and thereby are resistant to or less susceptible to a particular disorder.

The mammalian artificial megachromosome can be used to generate transgenic animals that stably express genes conferring desired traits,
15 such as genes conferring resistance to pathogenic viruses. Transgenic mice containing a transgene encoding an anti-HIV ribozyme provide a useful model for the development of stable transgenic animals using these methods.

20 1. Development of Control Transgenic Mice Expressing Anti-HIV Ribozyme

Control transgenic mice are generated in order to compare stability and amounts of transgene expression in mice developed using transgene DNA carried on a vector (control mice) with expression in mice developed using transgenes carried in an artificial megachromosome.

25 a. Development of Control Transgenic Mice Expressing β -galactosidase

One set of control transgenic mice was generated by microinjection of mouse embryos with the β -galactosidase gene alone. The microinjection procedure used to introduce the plasmid DNA into the
30 mouse embryos is as described in Example 13, but modified for use with embryos [see, e.g., Hogan et al. (1994) *Manipulating the Mouse Embryo*,

- A :Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, see, especially pages 255-264 and Appendix 3]. Fertilized mouse embryos [Strain CB6 obtained from Charles River Co.] were injected with 1 ng of plasmid pCH110 (Pharmacia) which had been
- 5 linearized by digestion with BamHI. This plasmid contains the β -galactosidase gene linked to the SV40 late promoter. The β -galactosidase gene product provides a readily detectable marker for successful transgene expression. Furthermore, these control mice provide confirmation of the microinjection procedure used to introduce the
 - 10 plasmid into the embryos. Additionally, because the megachromosome that is transferred to the mouse embryos in the model system (see below) also contains the β -galactosidase gene, the control transgenic mice that have been generated by injection of pCH110 into embryos serve as an analogous system for comparison of heterologous gene expression from a
 - 15 plasmid versus from a gene carried on an artificial chromosome.

After injection, the embryos are cultured in modified HTF medium under 5% CO₂ at 37°C for one day until they divide to form two cells. The two-cell embryos are then implanted into surrogate mother female mice [for procedures see, Manipulating the Mouse Embryo, A Laboratory

- 20 Manual (1994) Hogan et al., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 127 et seq.].

b. **Development of Control Transgenic Mice Expressing Anti-HIV Ribozyme**

- One set of anti-HIV ribozyme gene-containing control transgenic
- 25 mice was generated by microinjection of mouse embryos with plasmid pCEPUR-132 which contains three different genes: (1) DNA encoding an anti-HIV ribozyme, (2) the puromycin-resistance gene and (3) the hygromycin-resistance gene. Plasmid pCEPUR-132 was constructed by ligating portions of plasmid pCEP-132 containing the anti-HIV ribozyme
 - 30 gene (referred to as ribozyme D by Chang et al. [(1990) Clin. Biotech. 2:23-31]; see also U.S. Patent No. 5,144,019 to Rossi et al., particularly

Figure 4 of the patent) and the hygromycin-resistance gene with a portion of plasmid pCEPUR containing the puromycin-resistance gene.

Plasmid pCEP-132 was constructed as follows. Vector pCEP4 (Invitrogen, San Diego, CA; see also Yates *et al.* (1985) *Nature* 313:812-815) was digested with XhoI which cleaves in the multiple cloning site region of the vector. This ~10.4-kb vector contains the hygromycin-resistance gene linked to the thymidine kinase gene promoter and polyadenylation signal, as well as the ampicillin-resistance gene and ColE1 origin of replication and EBNA-1 (Epstein-Barr virus nuclear antigen) genes and OriP. The multiple cloning site is flanked by the cytomegalovirus promoter and SV40 polyadenylation signal.

XhoI-digested pCEP4 was ligated with a fragment obtained by digestion of plasmid 132 (see Example 4 for a description of this plasmid) with XhoI and Sall. This XhoI/Sall fragment contains the anti-HIV ribozyme gene linked at the 3' end to the SV40 polyadenylation signal. The plasmid resulting from this ligation was designated pCEP-132. Thus, in effect, pCEP-132 comprises pCEP4 with the anti-HIV ribozyme gene and SV40 polyadenylation signal inserted in the multiple cloning site for CMV promoter-driven expression of the anti-HIV ribozyme gene.

To generate pCEPUR-132, pCEP-132 was ligated with a fragment of pCEPUR. pCEPUR was prepared by ligating a 7.7-kb fragment generated upon NheI/NruI digestion of pCEP4 with a 1.1-kb NheI/SnaBI fragment of pBabe [see Morgenstern and Land (1990) *Nucleic Acids Res.* 18:3587-3596 for a description of pBabe] that contains the puromycin-resistance gene linked at the 5' end to the SV40 promoter. Thus, pCEPUR is made up of the ampicillin-resistance and EBNA1 genes, as well as the ColE1 and OriP elements from pCEP4 and the puromycin-resistance gene from pBabe. The puromycin-resistance gene in pCEPUR is flanked by the SV40 promoter (from pBabe) at the 5' end and the SV40 polyadenylation signal (from pCEP4) at the 3' end.

Plasmid pCEPUR was digested with XhoI and Sall and the fragment containing the puromycin-resistance gene linked at the 5' end to the SV40 promoter was ligated with XhoI-digested pCEP-132 to yield the ~12.1-kb plasmid designated pCEPUR-132. Thus, pCEPUR-132, in effect, comprises pCEP-132 with puromycin-resistance gene and SV40 promoter inserted at the XhoI site. The main elements of pCEPUR-132 are the hygromycin-resistance gene linked to the thymidine kinase promoter and polyadenylation signal, the anti-HIV ribozyme gene linked to the CMV promoter and SV40 polyadenylation signal, and the puromycin-resistance gene linked to the SV40 promoter and polyadenylation signal. The plasmid also contains the ampicillin-resistance and EBNA1 genes and the ColE1 origin of replication and OriP.

Zygotes were prepared from (C57BL/6JxCBA/J) F1 female mice [see, e.g., Manipulating the Mouse Embryo, A Laboratory Manual (1994) Hogan et al., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p. 429], which had been previously mated with a (C57BL/6JxCBA/J) F1 male. The male pronuclei of these F2 zygotes were injected [see, Manipulating the Mouse Embryo, A Laboratory Manual (1994) Hogan et al., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY] with pCEPUR-132 (~3 µg/ml), which had been linearized by digestion with NruI. The injected eggs were then implanted in surrogate mother female mice for development into transgenic offspring.

These primary carrier offspring were analyzed (as described below) for the presence of the transgene in DNA isolated from tail cells. Seven carrier mice that contained transgenes in their tail cells (but that may not carry the transgene in all their cells, i.e., they may be chimeric) were allowed to mate to produce non-chimeric or germ-line heterozygotes. The heterozygotes were, in turn, crossed to generate homozygote transgenic offspring.

Fertilized mouse embryos are microinjected (as described above)

isolated from fusion cell line G3D5* or H1D3* (described above). The megachromosomes are isolated as described herein. Megachromosomes isolated from either cell line carry the anti-HIV ribozyme (ribozyme D) gene as well as the hygromycin-resistance and β -galactosidase genes. The injected embryos are then developed into transgenic mice as

Alternatively, the megachromosome-containing cell line G3D5* or H1D3* is fused with mouse embryonic stem cells [see, e.g., U.S. Patent No. 5,453,357, commercially available; see Manipulating the Mouse Embryo, A Laboratory Manual (1994) Hogan *et al.*, eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pages 253-289] following standard procedures see also, e.g., Guide to Techniques in Mouse Development in Methods in Enzymology Vol. 25, Wassarman and De Pamphilis, eds. (1993), pages 803-932]. (It is also possible to deliver isolated megachromosomes into embryonic stem cells using the Microcell procedure [such as that described above].) The stem cells are cultured in the presence of a fibroblast [e.g., STO fibroblasts that are resistant to hygromycin and puromycin]. Cells of the resultant fusion cell line, which contains megachromosomes carrying the transgenes [i.e., anti-HIV ribozyme, hygromycin-resistance and β -galactosidase genes], are then transplanted into mouse blastocysts, which are in turn implanted into a surrogate mother female mouse where development into a transgenic mouse will occur.

30 thus may not be expressed in all cells.

3. Analysis of Transgenic Mice for Transgene Expression

Beginning when the transgenic mice, generated as described above, are three-to-four weeks old, they can be analyzed for stable expression of the transgenes that were transferred into the embryos [or
 5 fertilized eggs] from which they develop. The transgenic mice may be analyzed in several ways as follows.

a. Analysis of Cells Obtained from the Transgenic Mice

Cell samples [e.g., spleen cells, lymphocytes, tail cells] are
 10 obtained from the transgenic mice. Any cells may be tested for transgene expression. If, however, the mice are chimeras generated by microinjection of fertilized eggs with fusions of embryonic stem cells with megachromosome-containing cells, only cells from areas of the mouse that carry the transgene are expected to express the transgene. If the
 15 cells survive growth on hygromycin [or hygromycin and puromycin or neomycin, if the cells are obtained from mice generated by transfer of both antibiotic-resistance genes], this is one indication that they are stably expressing the transgenes. RNA isolated from the cells according to standard methods may also be analyzed by northern blot procedures to
 20 determine if the cells express transcripts that hybridize to nucleic acid probes based on the antibiotic-resistance genes.

Additionally, cells obtained from the transgenic mice may also be analyzed for β -galactosidase expression using standard assays for this marker enzyme [for example, by direct staining of the product of a
 25 reaction involving β -galactosidase and the X-gal substrate, see, e.g., Jones (1986) EMBO 5:3133-3142, or by measurement of β -galactosidase activity, see, e.g., Miller (1972) in Experiments in Molecular Genetics pp. 352-355, Cold Spring Harbor Press]. Analysis of β -galactosidase expression is particularly used to evaluate transgene expression in cells
 30 obtained from control transgenic mice in which the only transgene transferred into the embryo was the β -galactosidase gene.

Stable expression of the anti-HIV ribozyme gene in cells obtained from the transgenic mice may be evaluated in several ways. First, DNA isolated from the cells according to standard procedures may be subjected to nucleic acid amplification using primers corresponding to the ribozyme gene sequence. If the gene is contained within the cells, an amplified product of pre-determined size is detected upon hybridization of the reaction mixture to a nucleic acid probe based on the ribozyme gene sequence. Furthermore, DNA isolated from the cells may be analyzed using Southern blot methods for hybridization to such a nucleic acid probe. Second, RNA isolated from the cells may be subjected to northern blot hybridization to determine if the cells express RNA that hybridizes to nucleic acid probes based on the ribozyme gene. Third, the cells may be analyzed for the presence of anti-HIV ribozyme activity as described, for example, in Chang *et al.* (1990) *Clin. Biotech.* 2:23-31. In this analysis, RNA isolated from the cells is mixed with radioactively labeled HIV *gag* target RNA which can be obtained by *in vitro* transcription of *gag* gene template under reaction conditions favorable to *in vitro* cleavage of the *gag* target, such as those described in Chang *et al.* (1990) *Clin. Biotech.* 2:23-31. After the reaction has been stopped, the mixture is analyzed by gel electrophoresis to determine if cleavage products smaller in size than the whole template are detected; presence of such cleavage fragments is indicative of the presence of stably expressed ribozyme.

b. Analysis of Whole Transgenic Mice

Whole transgenic mice that have been generated by transfer of the anti-HIV ribozyme gene [as well as selection and marker genes] into embryos or fertilized eggs can additionally be analyzed for transgene expression by challenging the mice with infection with HIV. It is possible for mice to be infected with HIV upon intraperitoneal injection with high-producing HIV-infected U937 cells [see, e.g., Locardi *et al.* (1992) *J. Virol.* 66:1649-1654]. Successful infection may be confirmed by analysis

of DNA isolated from cells, such as peripheral blood mononuclear cells, obtained from transgenic mice that have been injected with HIV-infected human cells. The DNA of infected transgenic mice cells will contain HIV-specific gag and env sequences, as demonstrated by, for example,

- 5 nucleic acid amplification using HIV-specific primers. If the cells also stably express anti-HIV ribozyme, then analysis of RNA extracts of the cells should reveal the smaller gag fragments arising by cleavage of the gag transcript by the ribozyme.

- Additionally, the transgenic mice carrying the anti-HIV ribozyme
- 10 gene can be crossed with transgenic mice expressing human CD4 (i.e., the cellular receptor for HIV) [see Gillespie et al. (1993) Mol. Cell. Biol. 13:2952-2958; Hanna et al. (1994) Mol. Cell. Biol. 14:1084-1094; and Yeung et al. (1994) J. Exp. Med. 180:1911-1920, for a description of transgenic mice expressing human CD4]. The offspring of these crossed
- 15 transgenic mice expressing both the CD4 and anti-HIV ribozyme transgenes should be more resistant to infection [as a result of a reduction in the levels of active HIV in the cells] than mice expressing CD4 alone [without expressing anti-HIV ribozyme].

- 20 Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Hadlaczky, Gyula
Szalay, Aladar

(ii) TITLE OF THE INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES

(iii) NUMBER OF SEQUENCES: 12

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Brown, Martin, Haller & McClain
(B) STREET: 1660 Union Street
(C) CITY: San Diego
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 92101-2926

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 1.5

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE: 10-APR-1996
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Seidman, Stephanie L
(B) REGISTRATION NUMBER: 33,779
(C) REFERENCE/DOCKET NUMBER: 6869-402

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 619-238-0999
(B) TELEFAX: 619-238-0062
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1293 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAATTCATCA	TTTTTCANGT	CCTCAAGTGG	ATGTTTCTCA	TTTNCCATGA	TTTTAAGTTT	60
TCTCGCCATA	TTCTGGTCC	TACAGTGTGC	ATTTCTCCAT	TTTNCACGTT	TTNCAGTGAT	120
TTCGTCATTT	TCAAGTCCTC	AAGTGGATGT	TTCTCATTTN	CCATGAATTT	CAGTTTTCTN	180
GCCATATTCC	ACGTCCTACA	GNGGACATTT	CTAAATTTNC	CACCTTTTTT	AGTTTTCTCTC	240

GCCATATTTTC	ACGTCCTAAA	ATGTGTATTT	CTCGTTTNCC	GTGATTTTCA	GTTTTCTCGC	300
CAGATTCCAG	GTCCTATAAT	GTGCATTTCT	CATTTNNCAC	GTTTTTCAGT	GATTTTCGTCA	360
TTTTTTCAAG	TCGGCAAGTG	GATGTTTCTC	ATTTNCCATG	ATTTNCAGTT	TTCTTGNAAT	420
ATTCATGTC	CTACAATGAT	CATTTTTAAT	TTTCCACCTT	TTTTCATTTTC	CACGCCATAT	480
TTTCATGTCCT	AAAGTGTATA	TTTCTCCTTT	TCCGCGATT	TCAGTTTTCT	CGCCATATTC	540
CAGGTCCCTAC	AGTGTGCATT	CCTCATTTTT	CACCTTTTTTC	ACTGATTTTCG	TCATTTTTC	600
AGTCGTCAAC	TGGATCTTTC	TAATTTTCCA	TGATTTTCAG	TTATCTTGTC	ATATTCCATG	660
TCCTACAGTG	GACATTTCTA	AATTTTCCAA	CTTTTTC	TTTTCTCGAC	ATATTGACG	720
TGCTAAAGTG	TGTATTTCTT	ATTTTCCGTG	ATTTTCAGTT	TTCTCGCCAT	ATTCAGGTC	780
CTAATAGTGT	GCATTTCTCA	TTTTTCACGT	TTTTTCAGTGA	TTTCGTGATT	TTTTCCAGTT	840
GTCAAGGGGA	TGTTTCTCAT	TTTCCATGAG	TGTCAGTTTT	CTTGCTATAT	TCCATGTCCT	900
ACAGTGACAT	TTCTAAATAT	TATACCTTTT	TCAGTTTTTTC	TCACCATATT	TCACGTCCTA	960
AAGTATATAT	TTCTCATTTT	CCCTGATTTT	CAGTTTCCTT	GCCATATTCC	AGGTCCCTACA	1020
GTGTGCATTT	CTCATTTTTT	ACGTTTTTCA	GTAATTTCTT	CATTTTTTAA	GCCCTCAAAT	1080
GGATGTTTTCT	CATTTTCCAT	GATTTTTCAGT	TTTCTTGCCA	TATACCATGT	CCTACAGTGG	1140
ACATTTCTAA	ATTATCCACC	TTTTTCAGTT	TTTCATCGGC	ACATTTTCAGC	TCCTAAAGTG	1200
TGTATTTCTA	ATTTTCAGTG	ATTTTCAGTT	TTCTCGCCAT	ATTCCAGGAC	CTACAGTGTG	1260
CATTTCTCAT	TTTTTCACGTT	TTTCAGTGAA	TTC			1293

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1044 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTISENSE: NO
 (v) FRAGMENT TYPE:
 (vi) ORIGINAL SOURCE:
 (ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AGGCCTATGG	TGAAAAAGGA	AATATCTTCC	CCTGAAAACT	AGACAGAAGG	ATTCTCAGAA	60
TCITATTTGT	GATGTGCGCC	CCTCAACTAA	CAGTGTTGAA	GCTTTCTTTT	GATAGAGCAG	120
TTTTGAAACA	CTCTTTTGT	AAAATCTGCA	AGAGGATATT	TGGATAGCTT	TGAGGATTTT	180
CGTTGGAAAC	GGGATGTCT	TCATATAAAC	CCTAGACAGA	AGCATTCTCA	GAAGCTTCAT	240
TGGGATGTTT	CAGTTGAAGT	CACAGTGTTG	AACAGTCCCC	TTTCATAGAG	CAGGTTTGAA	300
ACACTCTTTT	TTGTAGTATC	TGGAAGTGGA	CATTTGGAGC	GATCTCAGGA	CTGCGGTGAA	360
AAAGGAAATA	TCTTCCAATA	AAAGCTAGAT	AGAGGCAATG	TCAGAAACCT	TTTTCATGAT	420
GTATCTACTC	AGCTAACAGA	GTTGAACCTT	CCTTTGAGAG	AGCAGTTTTG	AAACACTCTT	480
TTTGTGGAAT	CTGCAAGTGG	ATATTTGTCT	AGCTTTGAGG	ATTTTCGTTGG	GAAACGGGAT	540
TACATATAAA	AAGCAGACAG	CAGCATTCCC	AGAAACTTCT	TTGTGATGTT	TGCATTCAAG	600
TCACAGAGTT	GAACATTCCC	TTTCATAGAG	CAGGTTTGAA	ACACACTTTT	TGATGTATCT	660
GGATGTGGAC	ATTTGCAGCG	CTTTCAGGCC	TAAGGTGAAA	AGGAAATATC	TTCCCCTGAA	720
AAC TAGACAG	AAGCATTTCT	AGAAACTTAT	TTGTGATGTG	CGCCCTCAAC	TAACAGTGT	780
GAAGCTTTCT	TTTGATAGAG	GCAGTTTGA	AACACTCTTT	TGTGGAATCT	GCAAGTGGAT	840
ATTTGTCTAG	CTTTGAGGAT	TTCTTTGGAA	ACGGGATTAC	ATATAAAAAG	CAGACAGCAG	900
CATTCCGAGA	ATCTTGTTTG	TGATGTTTGC	ATTCAAGTCA	CAGAGTTGAA	CATTCCCTTT	960
CAGAGAGCAG	GTTTGAACAC	TCTTTTTATA	GTATCTGGAT	GTGGACATTT	GGAGCGCTTT	1020
CAGGGGGGAT	CCTCTAGAAT	TCCT				1044

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2492 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTISENSE: NO
 (v) FRAGMENT TYPE:
 (vi) ORIGINAL SOURCE:
 (ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CTGCAGCTGG	GGGTCTCCAA	TCAGGCAGGG	GCCCCCTACT	ACTCAGATGG	GGTGGCCGAG	60
TAGGGGAAGG	GGGTGCAGGC	TGCATGAGTG	GACACAGCTG	TAGGACTACC	TGGGGGCTGT	120
GGATCTATGG	GGGTGGGGAG	AAGCCCAGTG	ACAGTGCCTA	GAAGAGACAA	GGTGGCCTGA	180
GAGGGTCTGA	GGAACATAGA	GCTGGCCATG	TTGGGGCCAG	GTCTCAAGCA	GGAAGTGAGG	240
AATGGGACAG	GCCTTAGGAT	ACTCTACTCA	GTAGCCAGGA	TAGCAAGGAG	GGCTTGGGGT	300
TGCTATCCTG	GGGTTCAACC	CCCCAGTTG	AAGGCCCTGG	GGGAGATGGT	CCCAGGACAT	360
ATTACAATGG	ACACAGGAGG	TTGGGACACC	TGGAGTCACC	AAACAAAACC	ATGCCAAGAG	420
AGACCATGAG	TAGGGGTGTC	CAGTCCAGCC	CTCTGACTGA	GCTGCATTGT	TCAAATCCAA	480
AGGGCCCCCTG	CTGCCACCTA	GTGGCTGATG	GCATCCACAT	GACCCTGGGC	CACACGCGTT	540
TAGGGTCTCT	GTGAAGACCA	AGATCCTTGT	TACATTGAAC	GACTCCTAAA	TGAGCAGAGA	600
TTTCCACCTA	TTCGAAACAA	TCACATAAAA	TCCATCCTGG	AAAAAGCCTG	GGGGATGGCA	660
CTAAGGCTAG	GGATAGGGTG	GGATGAAGAT	TATAGTTACA	GTAAGGGGTT	TAGGGTTAGG	720
GATCAACGTT	GGTTAGGAGT	TAGGGATACA	TAGGGGTACC	GGTAGGGTTA	GGGGTTAGGG	780
TTAGGGGTTA	GGGTTAGGGT	TAGGGTTAGG	GTTAGGGTTA	GGGGTTAGGG	GTTAGGGTTA	840
GGGTTAGGTT	TTGGGGTGGC	GTATTTTGGT	CTTATACGCT	GTGTTCCACT	GGCAATGAAA	900
AGAGTTCCTG	TTTTTCCTTC	AGCAATTTGT	CATTTTTTAA	AGAGTTTAGC	AATTCTAACA	960
GATATAGACC	AGCTGTGCTA	TCTCATTTGT	GTTTTCAATT	GTAACCACAT	TGTGGTTTCA	1020
ATGTGTTTAC	TTGCCATCTG	TAGATCTTCT	TTGCGTGAGG	TGTCTGTTCA	GATGTGTGTG	1080
CATTTCTTGN	NTTNGGCTG	TTTAACTTAT	TGTTTAGTTT	TAATAATTTT	TTATATATTT	1140
GAAGACAAAT	CTTCTCAGA	TGTGTATTTG	CAAATATTTT	TTCAATATGA	GGCTTGCTTT	1200
TGTCTCTAAC	AAGGTCTCTT	CAGAGATAAC	TTAAATATAA	GAAATCCACA	CTGTCACTTC	1260
TTTTGTGTAT	ATCTACCTTT	TGTGTCAATT	GTTAAAATTC	ATTACCAAC	CCAAAGGCAG	1320
ATAGCTTTTC	TTCTATTGTT	TCTTCTAGAA	ATTTGTATAG	TTTTGCATTT	TTAGTGTAAG	1380
GATGATTTTG	AGTGATTATT	TGTGTAAGTT	GTAAAGTTTT	CGTCTATATC	CATATCATTT	1440
CTTATGGTTT	CCAATTAATC	GTTCCCTCAC	TATTTTTGGG	AAAGACACAG	GATAGTGGGC	1500
TTTGTTAGAG	TAGATAGGTA	GCTAGACATG	AACAGGAGGG	GGCCTCCTGG	AAAAGGGAAA	1560
GTCTGGGAAG	GCTCACCTGG	AGGACCACCA	AAAATTCACA	TATTAGTAGC	ATCTCTAGTG	1620
CTGGAGTGGA	TGGGCACTTG	TCAATTTGTG	GTAGGAGGGA	AAAGAGGTCC	TATGCAGAAA	1680
GAAACTCCCT	AGAACTCCTC	TGAAGATGCC	CCAATCATTC	ACTCTGCAAT	AAAAATGTCA	1740
GAATATTGCT	AGCTACATGC	TGATAAGGNN	AAAGGGGACA	TTCTTAAGTG	AAACCTGGCA	1800
CCATAAGTAC	AGATTAGGGC	AGAGAAGGAC	ATTCAAAAAGA	GGCAGGCGCA	GTAGGTACAA	1860
ACGTGATCGC	TGTCAGTGTG	CCTGGGATGG	CGGGAAGGAG	GCTGGTGCCA	GAGTGGATTC	1920
GTATTGATCA	CCACACATAT	ACCTCAACCA	ACAGTGAGGA	GGTCCCACAA	GCCTAAGTGG	1980
GGCAAGTTGG	GGAGCTAAGG	CAGTAGCAGG	AAAACCAGAC	AAAGAAAACA	GGTGGAGACT	2040
TGAGACAGAG	GCAGGAATGT	GAAGAAATCC	AAAATAAAAT	TCCCTGCACA	GGACTCTTAG	2100
GCTGTTTAAAT	GCATCGCTCA	GTCCCCTCC	TCCCTATTTT	TCTACAATAA	ACTCTTTTACA	2160
CTGTGTTTTCT	TTTCAATGAA	GTTATCTGCC	ATCTTTGTAT	TGCCTCTTGG	TGAAAATGTT	2220
TCTTCCAAGT	TAAACAAGAA	CTGGGACATC	AGCTCTCCCC	AGTAATAGCT	CCGTTTCAGT	2280
TTGAATTTTAC	AGAACTGATG	GGCTTAATAA	CTGGCGCTCT	GACTTTTAGTG	GTGCAGGAGG	2340
CCGTCAACAC	GGGACCAAGA	GTGCCCTGCC	TAGTCCCCAT	CTGCCCCGAG	GTGGCGGCTG	2400
CCTCGACACT	GACAGCAATA	GGGTCCGGCA	GTGTCCCCAG	CTGCCAGCAG	GGGGCGTACG	2460
ACGACTACAC	TGTGAGCAAG	AGGGCCCTGC	AG			2492

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 28 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

28

(i) SEQUENCE CHARACTERISTICS:

- ```
(ii) MOLECULE TYPE: Genomic DNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:
```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

29

(i) SEQUENCE CHARACTERISTICS:

- ```
(ii) MOLECULE TYPE: RNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:
```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCGCTTAATA CTCTGATGAG TCCGTGAGGA CGAAACGCTC TCGCACC

(i) SEQUENCE CHARACTERISTICS:

- ```
(ii) MOLECULE TYPE: Genomic DNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
```



## (x) PUBLICATION INFORMATION:

PATENT NO.: 5,418,155

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| AGC TTA AAG ATG ACT TCG AAA GTT TAT GAT CCA GAA CAA AGG AAA CGG | 48  |
| Ser Leu Lys Met Thr Ser Lys Val Tyr Asp Pro Glu Gln Arg Lys Arg |     |
| 1 5 10 15                                                       |     |
| ATG ATA ACT GGT CCG CAG TGG TGG GCC AGA TGT AAA CAA ATG AAT GTT | 96  |
| Met Ile Thr Gly Pro Gln Trp Trp Ala Arg Cys Lys Gln Met Asn Val |     |
| 20 25 30                                                        |     |
| CTT GAT TCA TTT ATT AAT TAT TAT GAT TCA GAA AAA CAT GCA GAA AAT | 144 |
| Leu Asp Ser Phe Ile Asn Tyr Tyr Asp Ser Glu Lys His Ala Glu Asn |     |
| 35 40 45                                                        |     |
| GCT GTT ATT TTT TTA CAT GGT AAC GCG GCC TCT TCT TAT TTA TGG CGA | 192 |
| Ala Val Ile Phe Leu His Gly Asn Ala Ala Ser Ser Tyr Leu Trp Arg |     |
| 50 55 60                                                        |     |
| CAT GTT GTG CCA CAT ATT GAG CCA GTA GCG CGG TGT ATT ATA CCA GAT | 240 |
| His Val Val Pro His Ile Glu Pro Val Ala Arg Cys Ile Ile Pro Asp |     |
| 65 70 75 80                                                     |     |
| CTT ATT GGT ATG GGC AAA TCA GGC AAA TCT GGT AAT GGT TCT TAT AGG | 288 |
| Leu Ile Gly Met Gly Lys Ser Gly Lys Ser Gly Asn Gly Ser Tyr Arg |     |
| 85 90 95                                                        |     |
| TTA CTT GAT CAT TAC AAA TAT CTT ACT GCA TGG TTG AAC TTC TTA ATT | 336 |
| Leu Leu Asp His Tyr Lys Tyr Leu Thr Ala Trp Leu Asn Phe Leu Ile |     |
| 100 105 110                                                     |     |
| TAC CAA AGA AGA TCA TTT TTT GTC GGC CAT GAT TGG GGT GCT TGT TTG | 384 |
| Tyr Gln Arg Arg Ser Phe Phe Val Gly His Asp Trp Gly Ala Cys Leu |     |
| 115 120 125                                                     |     |
| GCA TTT CAT TAT AGC TAT GAG CAT CAA GAT AAG ATC AAA GCA ATA GTT | 432 |
| Ala Phe His Tyr Ser Tyr Glu His Gln Asp Lys Ile Lys Ala Ile Val |     |
| 130 135 140                                                     |     |
| CAC GCT GAA AGT GTA GTA GAT GTG ATT GAA TCA TGG GAT GAA TGG CCT | 480 |
| His Ala Glu Ser Val Val Asp Val Ile Glu Ser Trp Asp Glu Trp Pro |     |
| 145 150 155 160                                                 |     |
| GAT ATT GAA GAA GAT ATT GCG TTG ATC AAA TCT GAA GAA GGA GAA AAA | 528 |
| Asp Ile Glu Glu Asp Ile Ala Leu Ile Lys Ser Glu Glu Gly Glu Lys |     |
| 165 170 175                                                     |     |
| ATG GTT TTG GAG AAT AAC TTC TTC GTG GAA ACC ATG TTG CCA TCA AAA | 576 |
| Met Val Leu Glu Asn Asn Phe Phe Val Glu Thr Met Leu Pro Ser Lys |     |
| 180 185 190                                                     |     |
| ATC ATG AGA AAG TTA GAA CCA GAA GAA TTT GCA GCA TAT CTT GAA CCA | 624 |
| Ile Met Arg Lys Leu Glu Pro Glu Glu Phe Ala Ala Tyr Leu Glu Pro |     |
| 195 200 205                                                     |     |
| TTC AAA GAG AAA GGT GAA GTT CGT CGT CCA ACA TTA TCA TGG CCT CGT | 672 |
| Phe Lys Glu Lys Gly Glu Val Arg Arg Pro Thr Leu Ser Trp Pro Arg |     |
| 210 215 220                                                     |     |

|                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |     |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-----|
| GAA<br>Glu<br>225 | ATC<br>Ile        | CCG<br>Pro        | TTA<br>Leu        | GTA<br>Val        | AAA<br>Lys<br>230 | GGT<br>Gly        | GGT<br>Gly        | AAA<br>Lys        | CCT<br>Pro        | GAC<br>Asp<br>235 | GTT<br>Val        | GTA<br>Val        | CAA<br>Gln        | ATT<br>Ile        | GTT<br>Val<br>240 | 720 |
| AGG<br>Arg        | AAT<br>Asn        | TAT<br>Tyr        | AAT<br>Asn        | GCT<br>Ala<br>245 | TAT<br>Tyr        | CTA<br>Leu        | CGT<br>Arg        | GCA<br>Ala        | AGT<br>Ser<br>250 | GAT<br>Asp        | GAT<br>Asp        | TTA<br>Leu        | CCA<br>Pro        | AAA<br>Lys<br>255 | ATG<br>Met        | 768 |
| TTT<br>Phe        | ATT<br>Ile        | GAA<br>Glu        | TCG<br>Ser<br>260 | GAT<br>Asp        | CCA<br>Pro        | GGA<br>Gly        | TTC<br>Phe        | TTT<br>Phe<br>265 | TCC<br>Ser        | AAT<br>Asn        | GCT<br>Ala        | ATT<br>Ile        | GTT<br>Val<br>270 | GAA<br>Glu        | GGC<br>Gly        | 816 |
| GCC<br>Ala        | AAG<br>Lys        | AAG<br>Lys<br>275 | TTT<br>Phe        | CCT<br>Pro        | AAT<br>Asn        | ACT<br>Thr        | GAA<br>Glu<br>280 | TTT<br>Phe        | GTC<br>Val        | AAA<br>Lys        | GTA<br>Val        | AAA<br>Lys<br>285 | GGT<br>Gly        | CTT<br>Leu        | CAT<br>His        | 864 |
| TTT<br>Phe        | TCG<br>Ser<br>290 | CAA<br>Gln        | GAA<br>Glu        | GAT<br>Asp        | GCA<br>Ala        | CCT<br>Pro<br>295 | GAT<br>Asp        | GAA<br>Glu        | ATG<br>Met        | GGA<br>Gly        | AAA<br>Lys<br>300 | TAT<br>Tyr        | ATC<br>Ile        | AAA<br>Lys        | TCG<br>Ser        | 912 |
| TTC<br>Phe<br>305 | GTT<br>Val        | GAG<br>Glu        | CGA<br>Arg        | GTT<br>Val        | CTC<br>Leu<br>310 | AAA<br>Lys        | AAT<br>Asn        | GAA<br>Glu        | CAA<br>Gln        | TAA               |                   |                   |                   |                   |                   | 945 |

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- ```
(ii) MOLECULE TYPE: Genomic DNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:
```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TTTGAATTC A TGTACAGGAT GCAACTCCTG

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- ```
(ii) MOLECULE TYPE: Genomic DNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:
```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TTTGAATTCA GTAGGTGCAC TGTTCGTCAC